Statistical Analysis of Subcellular Proteins in Microscopy Imagery

Raquel A. Romano

Lawrence Berkeley National Laboratory

romano@hpcrd.lbl.gov

Joint Work with

Bahram Parvin, Imaging and Informatics Group, LBNL

Mary Helen Barcellos-Hoff, Life Sciences Division, LBNL

March 5, 2005

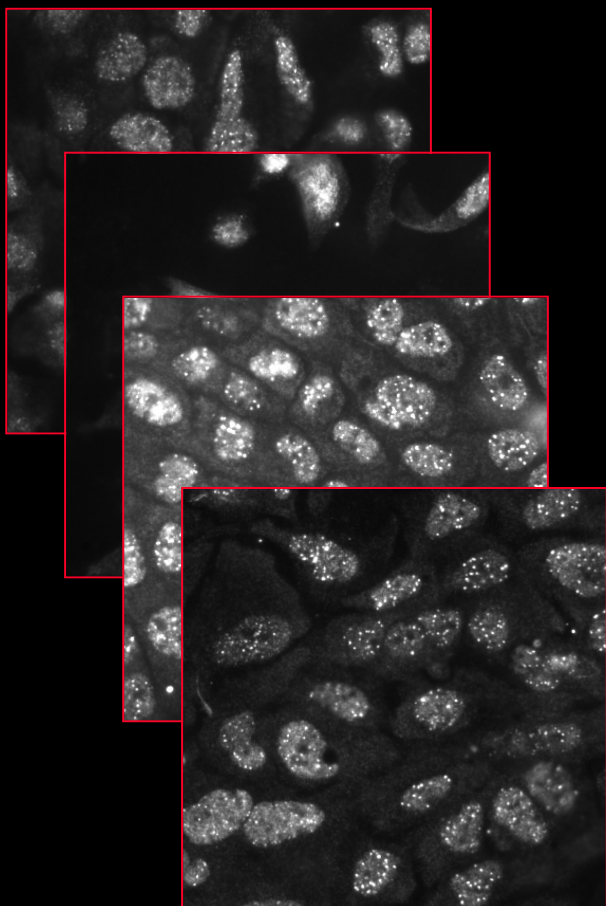
Motivating Applications: Radiation Biology

Problem: Quantify effects of radiation on cells by analyzing presence of proteins known to play crucial roles in DNA repair.

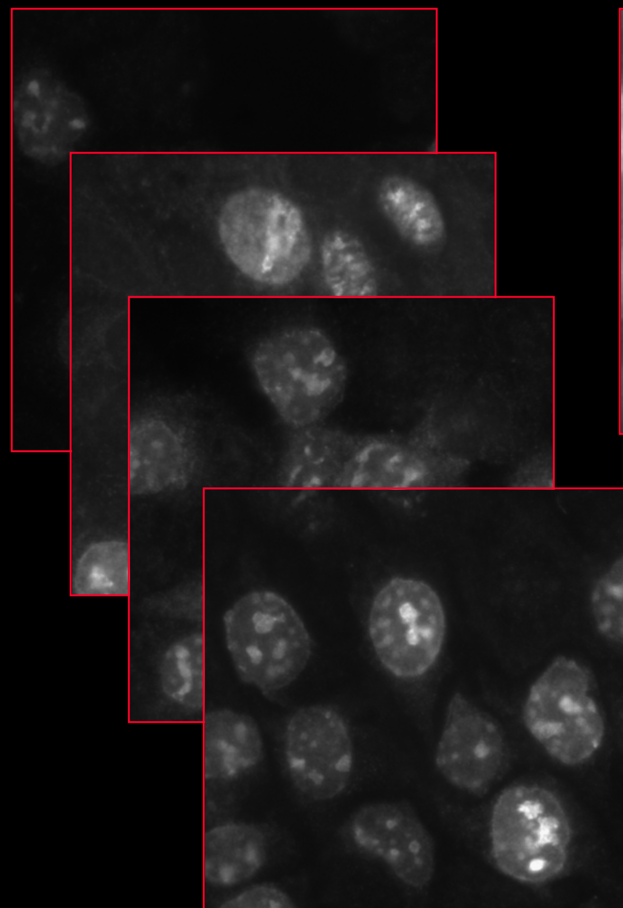
Data: Proteins in cell cultures are tagged with fluorescent probes, stained, irradiated, imaged.

Biological Challenge:
High Throughput Microscopy Analysis

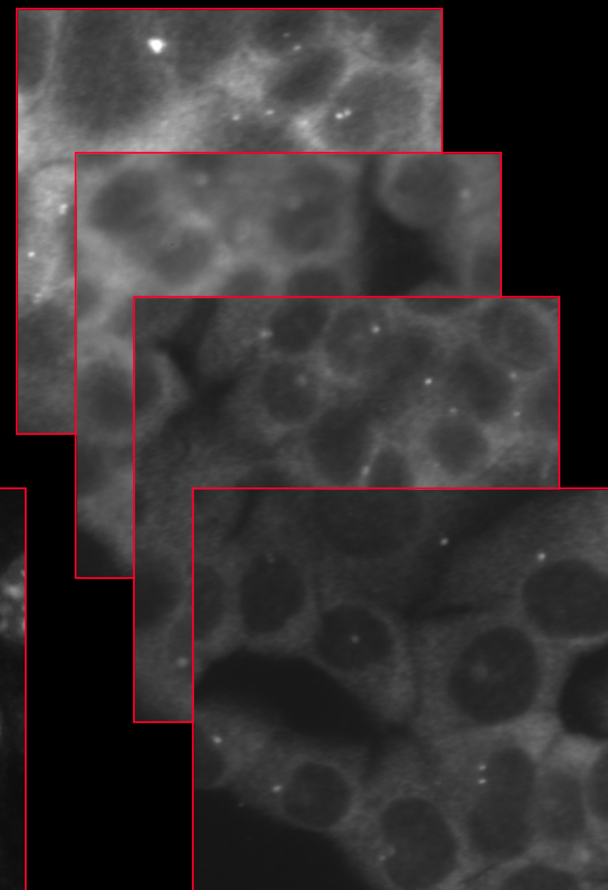
ATMp



Ki67



gammatubulin

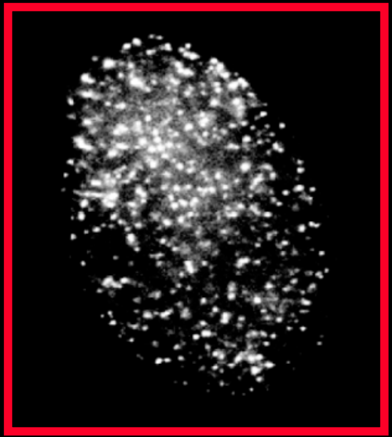


Sources of Image Variation

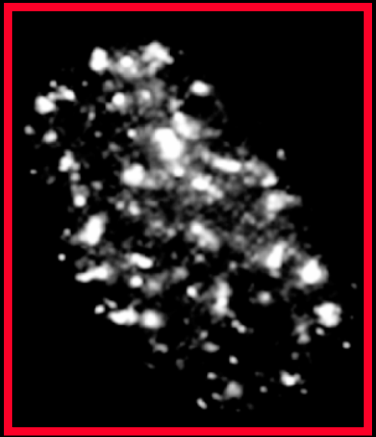
- Radiation dose/quality: high vs. low, Xray photons (sparse) vs. heavy charged particles (dense)
- Microscopy: fluorescence, confocal, deconvolution
- Time course
- Cell type and stage of mitotic cycle
- Direct vs. indirect effects of radiation
- Sample preparation

Sources of Image Variation

- Radiation dose/quality: high vs. low, Xray photons (sparse) vs. heavy charged particles (dense)
- Microscopy: fluorescence, confocal, deconvolution
- Time course
- Cell type and stage of mitotic cycle
- Direct vs. indirect effects of radiation
- Sample preparation



early foci



late foci

(Petrini & Stracker, 2003)

Mathematical Challenge

- **Current microscopy analysis:** easy to design ad hoc, heuristic methods to get quick scientific results.
- **Disadvantage:** special purpose tools are designed for each new biological research question.

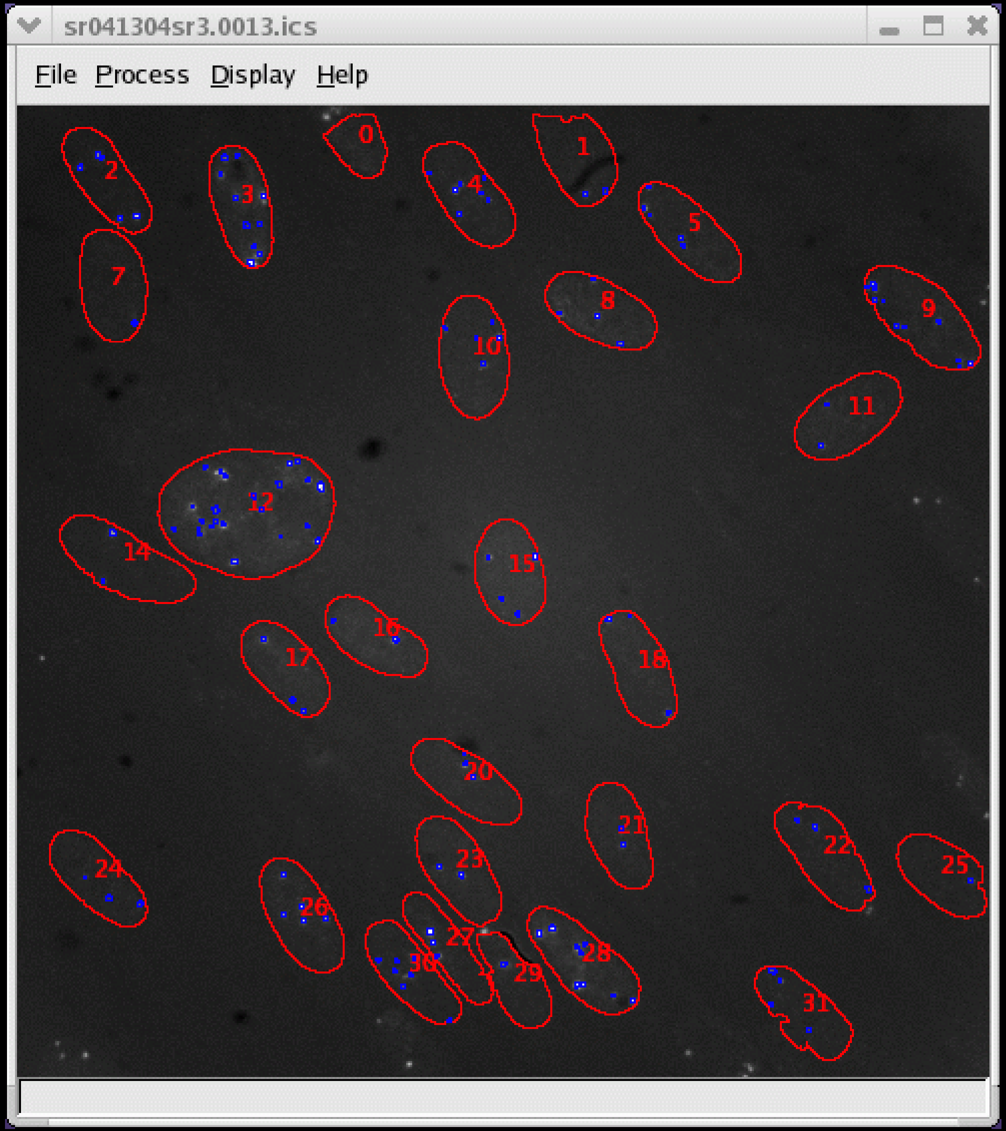
Mathematical Challenge

- **Current microscopy analysis:** easy to design ad hoc, heuristic methods to get quick scientific results.
- **Disadvantage:** special purpose tools are designed for each new biological research question.
- **Mathematical models:** can yield general solutions to a greater range of scientific questions and data.
- **Future microscopy analysis:** must move away from patchwork of image processing tools and toward principled statistical modeling and analysis.

Outline

1. Microscopy Analysis: Biological and Mathematical Challenges
2. Current Approach: Microscopy Image Analysis Application
3. Statistical Approach: Background
4. Statistical Approach: Preliminary Results
5. Current Directions

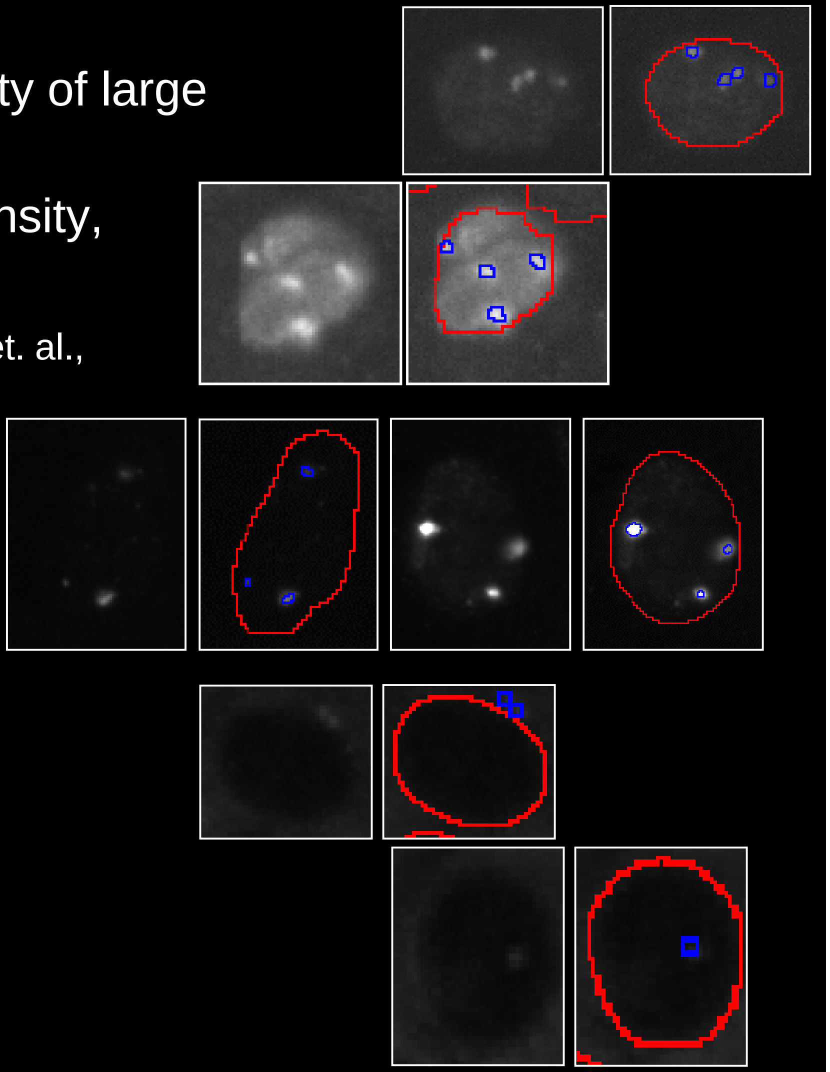
Foci Analysis Application



Foci Analysis Application

Pros

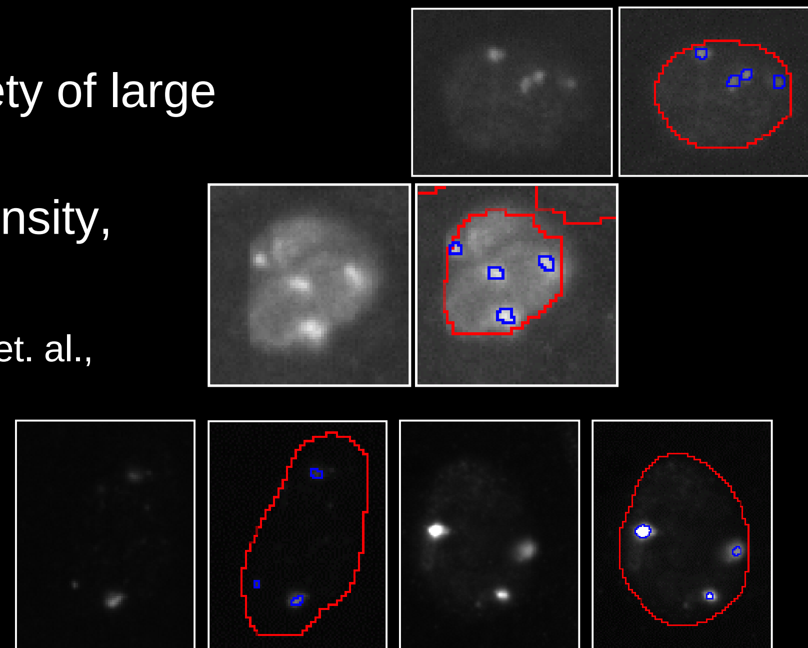
- Achieves high-accuracy foci detection for variety of large image sets
- Automatic collection of foci statistics (size, intensity, contrast, quantity per nucleus)
- Currently in use by radiation biologists (Costes, et. al., Bioastronautics Investigators' Workshop, 2005)



Foci Analysis Application

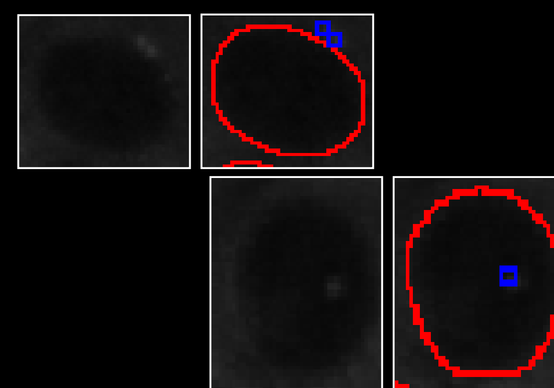
Pros

- Achieves high-accuracy foci detection for variety of large image sets
- Automatic collection of foci statistics (size, intensity, contrast, quantity per nucleus)
- Currently in use by radiation biologists (Costes, et. al., Bioastronautics Investigators' Workshop, 2005)

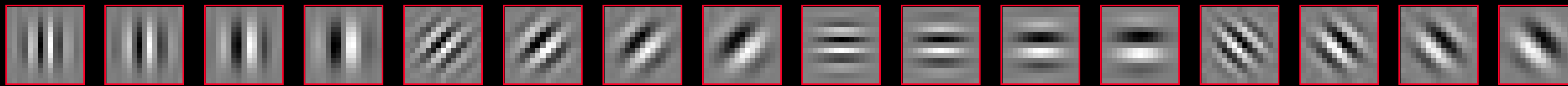


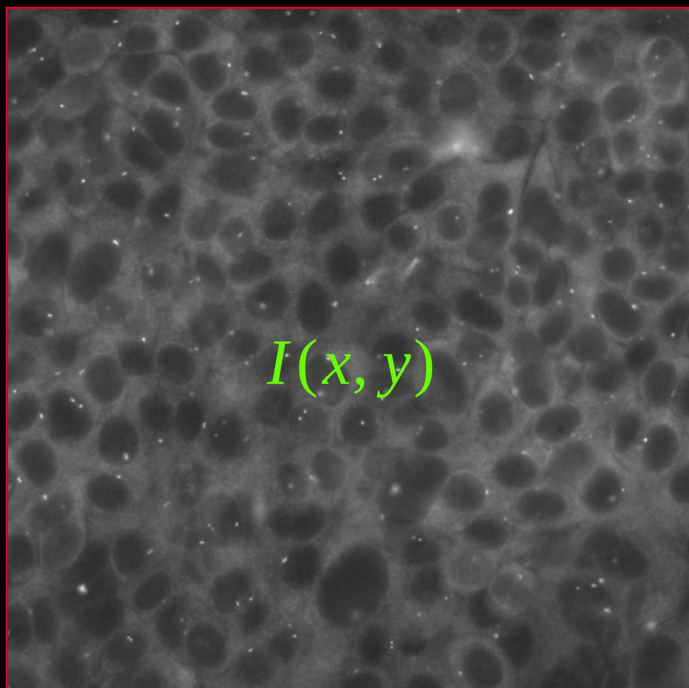
Cons

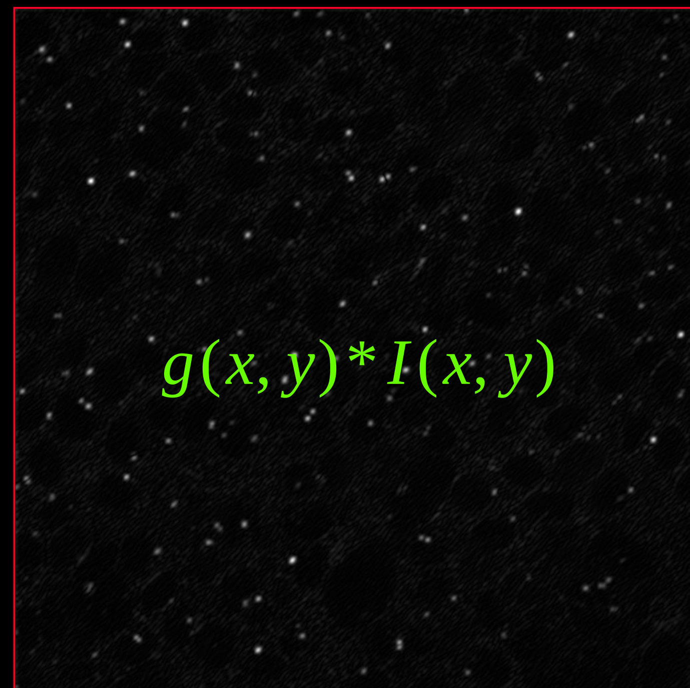
- Parameters adjusted by user can bias results
- Designed for a specific type of feature
- Toolbox approach not easily extendable to other tasks



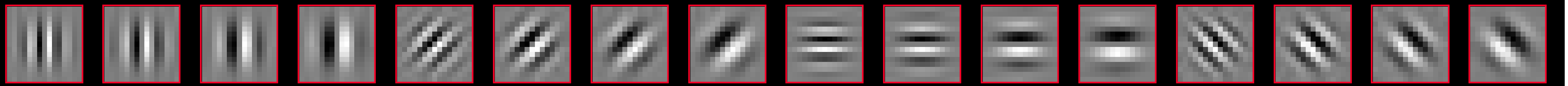
Feature Detection via Filter Banks

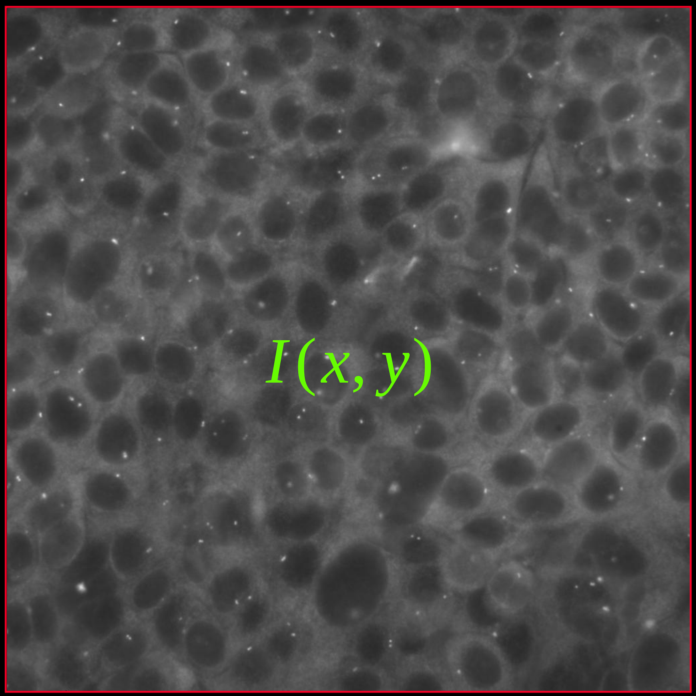




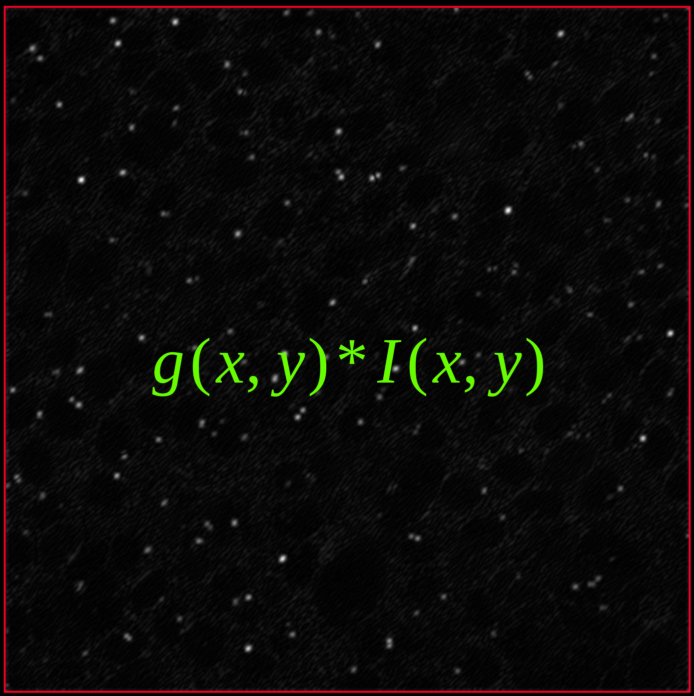


Feature Detection via Filter Banks





Gabor wavelet transform →



$$g(x,y,\theta,\lambda)=$$

$$\exp\left[-\frac{1}{2}\left\{\frac{(x\cos\theta+y\sin\theta)^2}{\sigma_x^2}+\frac{(-x\sin\theta+y\cos\theta)^2}{\sigma_y^2}\right\}\right]\exp\left[\frac{2\pi(x\cos\theta+y\sin\theta)}{\lambda}i\right]$$

Outline

1. Microscopy Analysis: Biological and Mathematical Challenges
2. Current Approach: Microscopy Image Analysis Application
3. Statistical Approach: Background
4. Statistical Approach: Preliminary Results
5. Current Directions

How to Learn Basis From Data?

Independent Components Analysis (Comon, 1991)

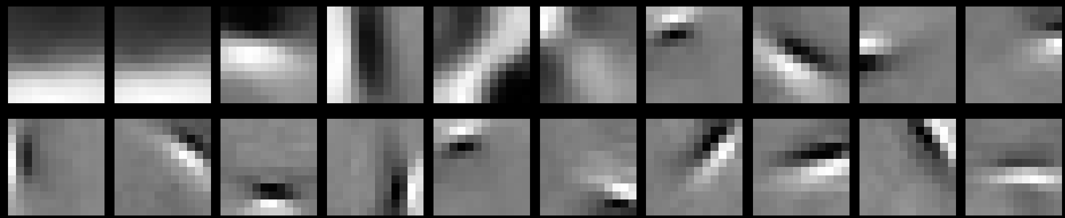
- Use statistical criteria to define optimal basis for decomposing images into combinations of basis functions
- Optimization yields multi-oriented, multiscale, bandpass basis functions specifically tuned to data set

How to Learn Basis From Data?

Independent Components Analysis (Comon, 1991)

- Use statistical criteria to define optimal basis for decomposing images into combinations of basis functions
- Optimization yields multi-oriented, multiscale, bandpass basis functions specifically tuned to data set

example basis functions



Sparse Coding (Olshausen-Field, 1996)
Independents Components Analysis (Bell & Sejnowski, 1997)

- More localized than Fourier filters , arbitrary frequencies and orientations, unlike Gabor and wavelet filters

Independent Components Analysis

- Consider an arbitrary NxN subimage of a sample image for fixed N:

Independent Components Analysis

- Consider an arbitrary NxN subimage of a sample image for fixed N: 

- Treat each subimage as an observed vector variable $\mathbf{x}$ that is a mixture of unknown basis functions $\mathbf{a}_i$ with unknown coefficients s_i

$$\mathbf{x} = s_1\mathbf{a}_1 + s_2\mathbf{a}_2 + \cdots + s_n\mathbf{a}_n$$

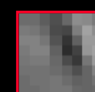
 $= s_1$  $+ s_2$  $+ \cdots + s_n$ 

Independent Components Analysis

- Consider an arbitrary NxN subimage of a sample image for fixed N:

- Treat each subimage as an observed vector variable $\mathbf{x}$ that is a mixture of unknown basis functions $\mathbf{a}_i$ with unknown coefficients s_i

$$\mathbf{x} = s_1\mathbf{a}_1 + s_2\mathbf{a}_2 + \cdots + s_n\mathbf{a}_n$$

 $= s_1$  $+ s_2$  $+ \cdots + s_n$ 

- Unknowns: basis functions $\mathbf{a}_i$ and coefficients s_i
- Given: only observations $\mathbf{x}$

$$\mathbf{x} = \mathbf{A}\mathbf{s}$$

Optimization

$$\mathbf{x} = \mathbf{A}\mathbf{s}$$

- Express unknown coefficients in terms of inverse $\mathbf{W}$ of unknown mixing matrix $\mathbf{A}$

$$\mathbf{s}^* = \mathbf{W}\mathbf{x}$$

- Coefficient estimates $\mathbf{s}^*$ are called *independent components*

Optimization

$$\mathbf{x} = \mathbf{A}\mathbf{s}$$

- Express unknown coefficients in terms of inverse $\mathbf{W}$ of unknown mixing matrix $\mathbf{A}$

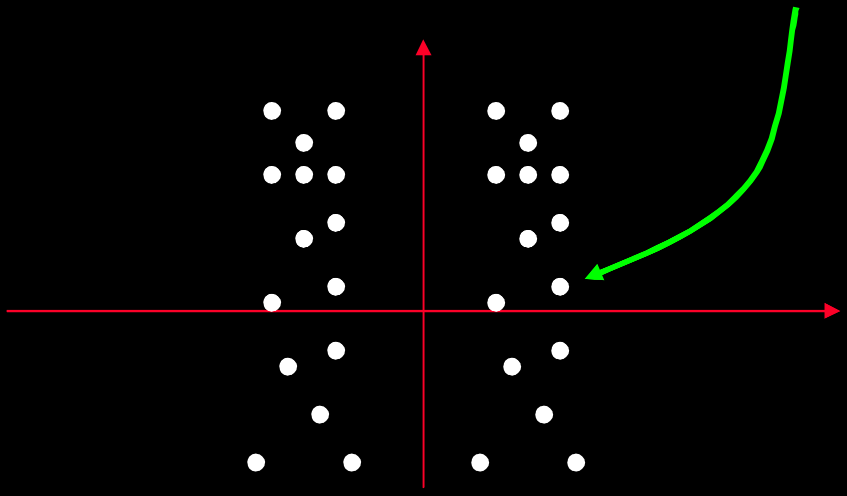
$$\mathbf{s}^* = \mathbf{W}\mathbf{x}$$

- Coefficient estimates $\mathbf{s}^*$ are called *independent components*
- Optimize a function $J(\mathbf{s}^*)$ that encourages the independent components to be as *non-Gaussian* as possible

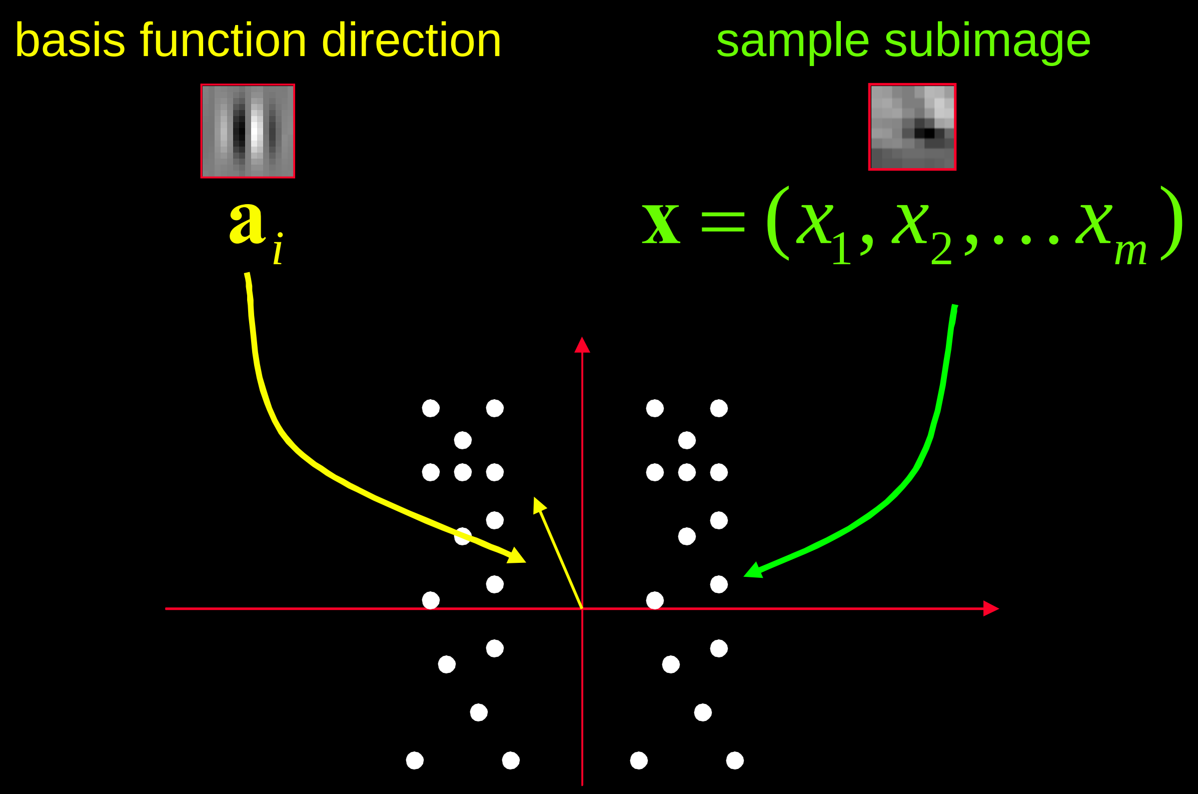
Non-Gaussianity: What and Why?

sample subimage

$$\mathbf{x} = (x_1, x_2, \ldots x_m)$$

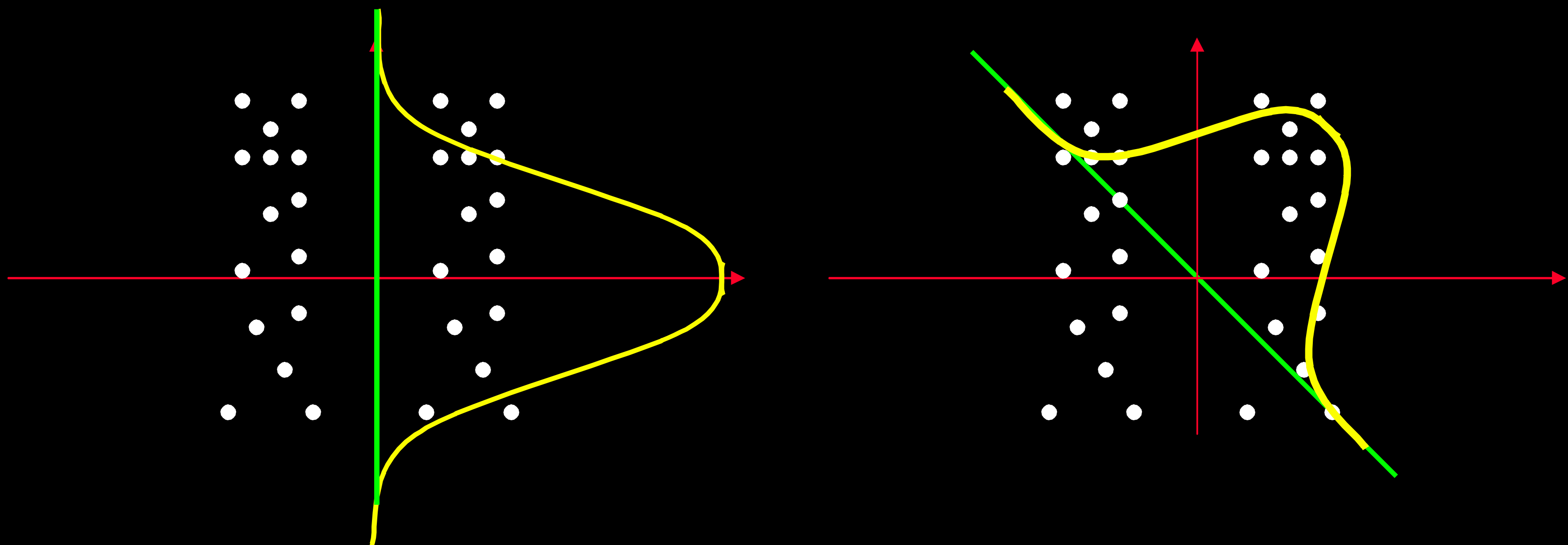


Non-Gaussianity: What and Why?



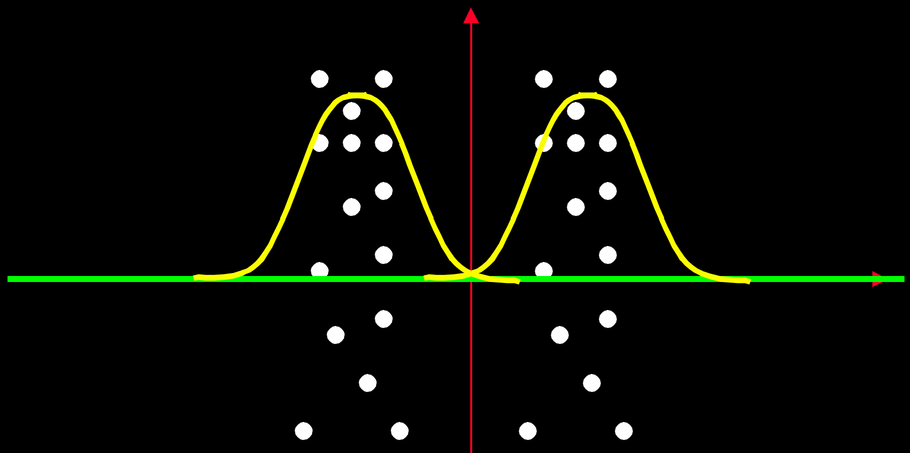
Non-Gaussianity: What and Why?

- Projection pursuit: random subspace projections of structured point clouds are typically Gaussian



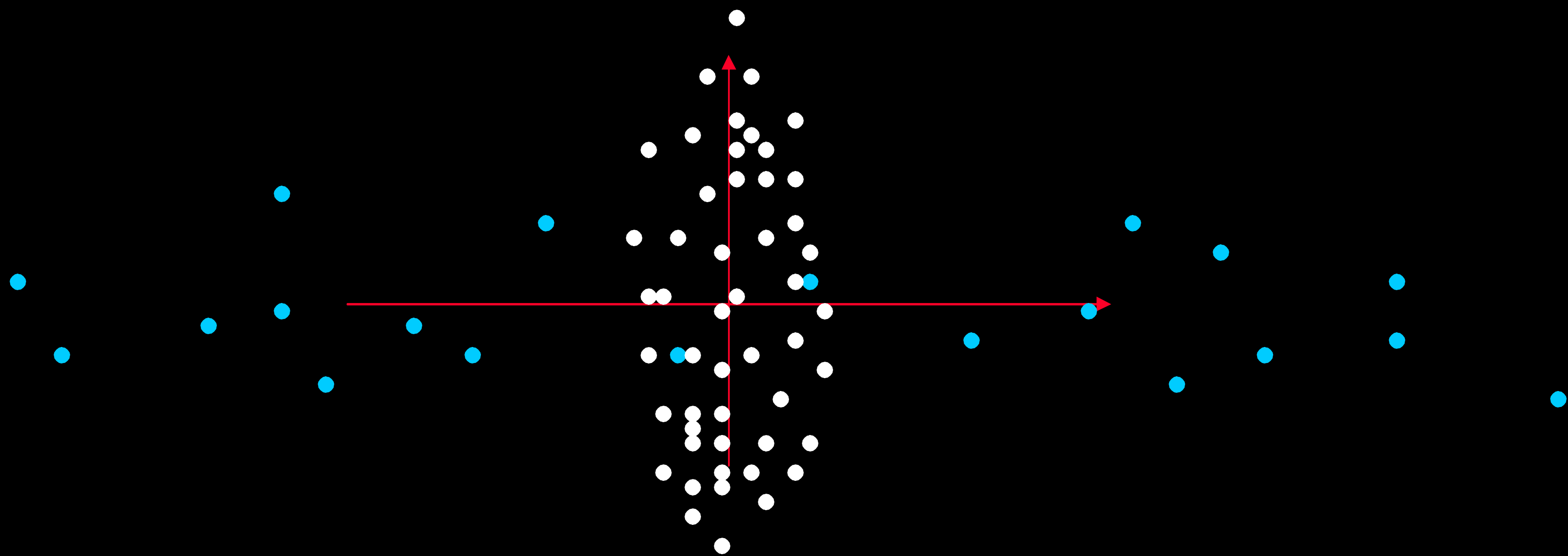
Non-Gaussianity: What and Why?

- Projection pursuit: random subspace projections of structured point clouds are typically Gaussian
- *Interesting* projections are not Gaussian, e.g., mixture of Gaussians for cluster discrimination



Non-Gaussianity: What and Why?

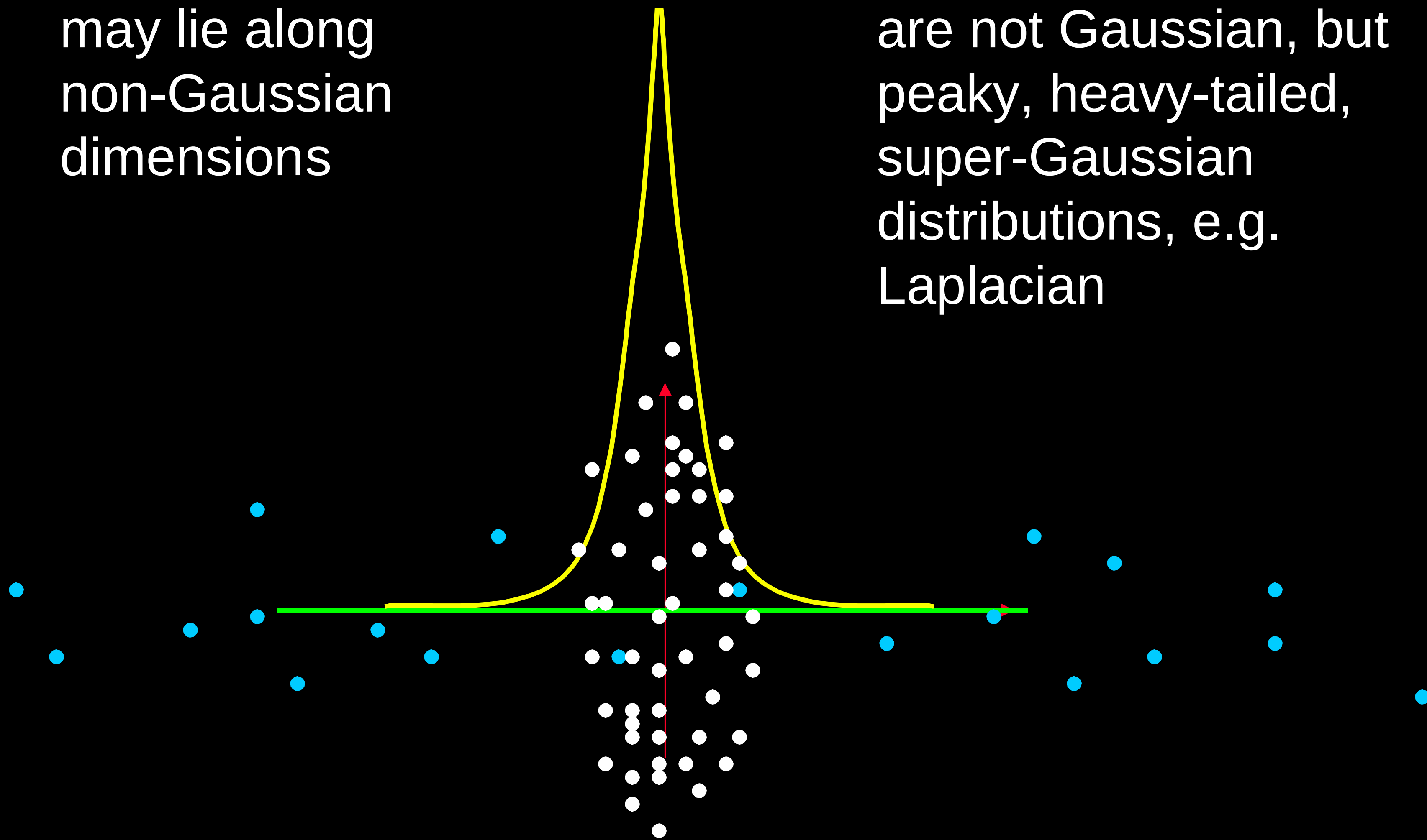
- Salient features may lie along non-Gaussian dimensions



Non-Gaussianity: What and Why?

- Salient features may lie along non-Gaussian dimensions

- Interesting directions are not Gaussian, but peaky, heavy-tailed, super-Gaussian distributions, e.g. Laplacian



Recall: Optimization Constraints

$$\mathbf{x} = \mathbf{A}\mathbf{s}$$

- Optimize a function $J(\mathbf{s}^*)$ that encourages the independent components to be as *non-Gaussian* as possible

- Output: basis functions and independent components

$$\mathbf{x} = s_1 \mathbf{a}_1 + s_2 \mathbf{a}_2 + \cdots + s_n \mathbf{a}_n$$

Non-Gaussianity: Contrast Functions

Entropy: measure of how random vs. structured a random variable is

$$H(\mathbf{y}) = -\int f(\mathbf{y}) \log f(\mathbf{y}) d\mathbf{y}$$

Negentropy: difference in entropy w.r.t. Gaussian with same variance

$$J(\mathbf{y}) = H(\mathbf{y}_{\text{gauss}}) - H(\mathbf{y})$$

Non-Gaussianity: Contrast Functions

Entropy: measure of how random vs. structured a random variable is

$$H(\mathbf{y}) = -\int f(\mathbf{y}) \log f(\mathbf{y}) d\mathbf{y}$$

Negentropy: difference in entropy w.r.t. Gaussian with same variance

$$J(\mathbf{y}) = H(\mathbf{y}_{\text{gauss}}) - H(\mathbf{y})$$

Negentropy Approximation:
c constant
$\mathbf{v}$ standard Gaussian variable
G non-quadratic function

$$J(\mathbf{y}) \approx c[E\{G(\mathbf{y})\} - E\{G(\mathbf{v})\}]^2$$

FastICA: A. Hyvärinen. Fast and Robust Fixed-Point Algorithms for Independent Component Analysis. *IEEE Transactions on Neural Networks* 1999.

Outline

1. Microscopy Analysis: Biological and Mathematical Challenges

2. Current Approach: Microscopy Image Analysis Application

3. Statistical Approach: Background

4. Statistical Approach: Preliminary Results

5. Current Directions

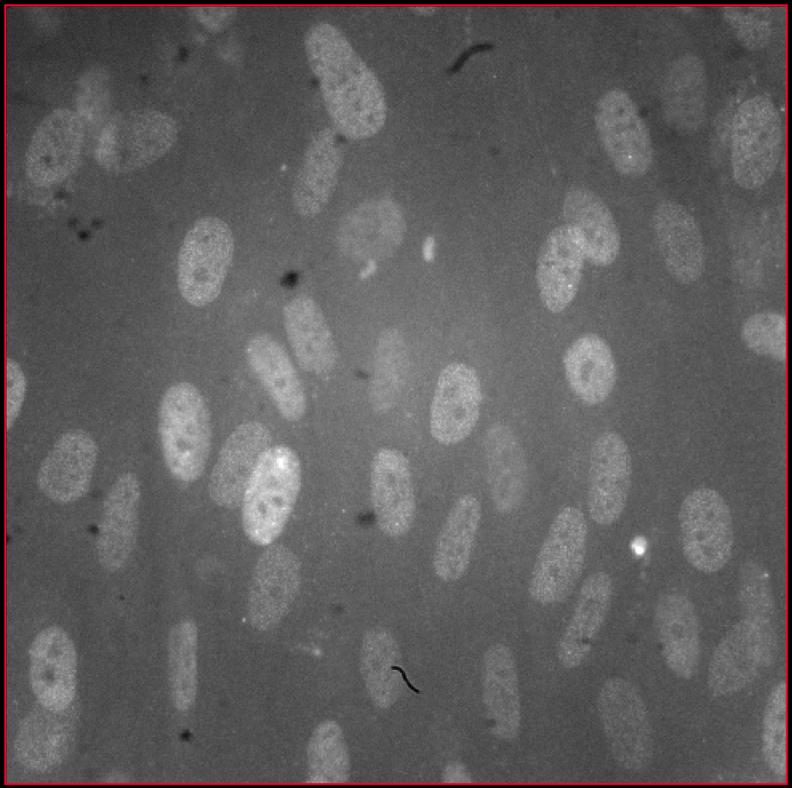
Independent Components of Microscopy Imagery

Goals

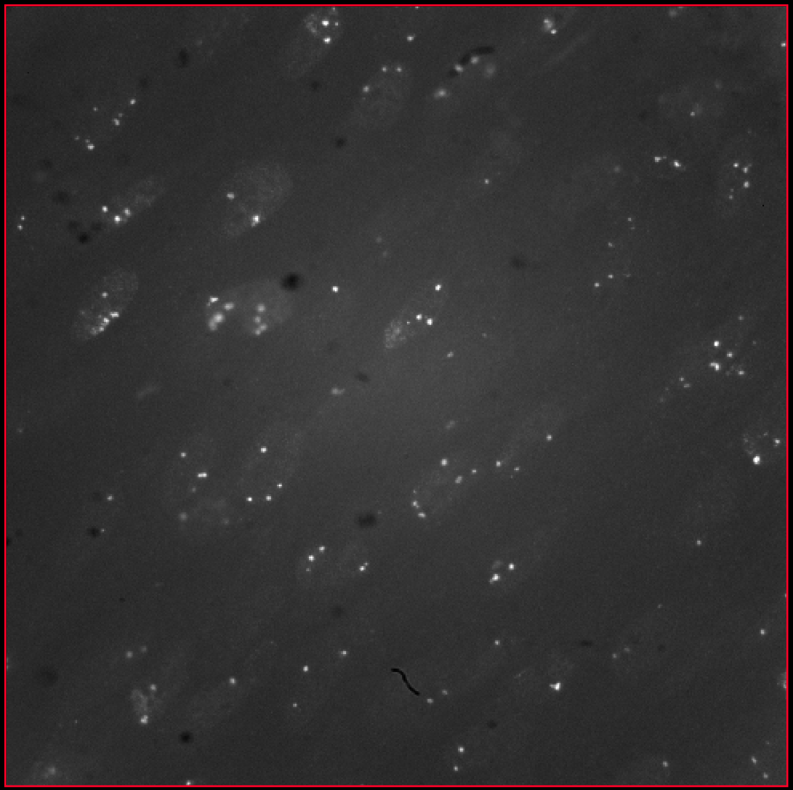
- Represent salient features characterizing cell differences, e.g. foci
- Distinguish differently treated cells and images
- Design task-specific algorithms: feature extraction, cell and image classification

Preliminary Test

Control Group



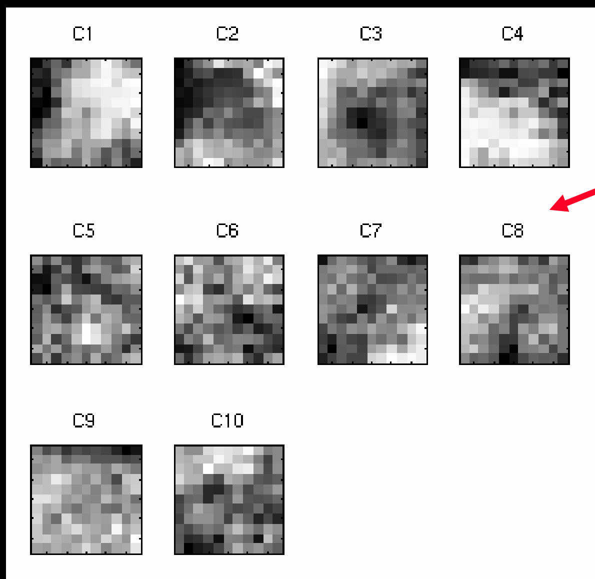
Dose: 30cGy, Time: 10min

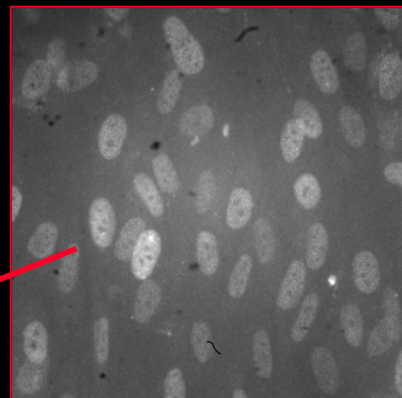


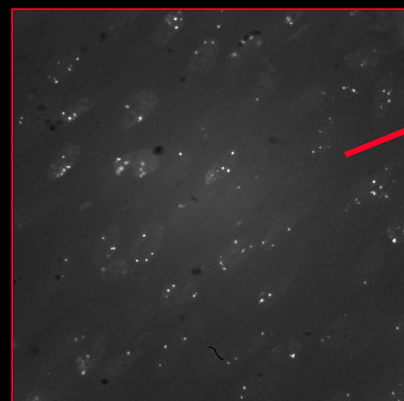
- *Input data: 15x15 pixel subimages within nuclei*
- *Centered at local maxima*
- *Small training set (~1000 subimages each)*

Example Microscopy Basis Functions

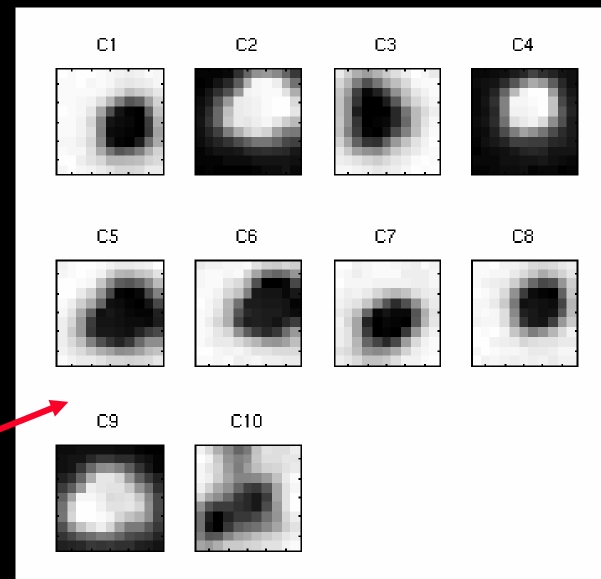
Basis Functions
Control Group







Basis Functions
Dose: 30cGy, Time: 10min



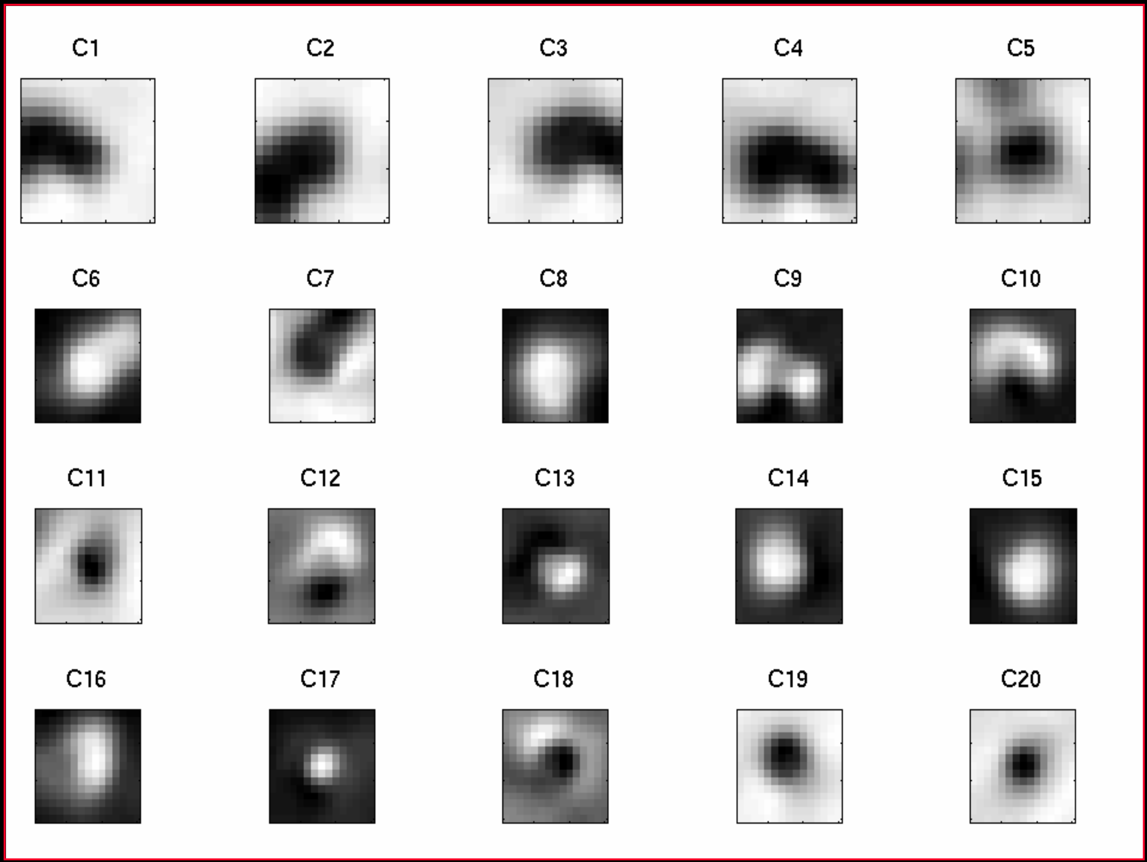
- *Subimages centered and whitened (decorrelated, unit variance)*
- *Data dimensionality may be reduced by PCA*
- *Basis functions defined up to unknown sign ambiguity*

Large Data Set

- *~100 1.3 Mb images*
- *9 classes (varying dose/time treatment groups)*
- *~10 images per class*
  2 training, 8 testing
- *~10,000 15x15 subimages*
- *random sample of ~2000 training examples*
- *dimensionality reduction by PCA from 225 to 20*

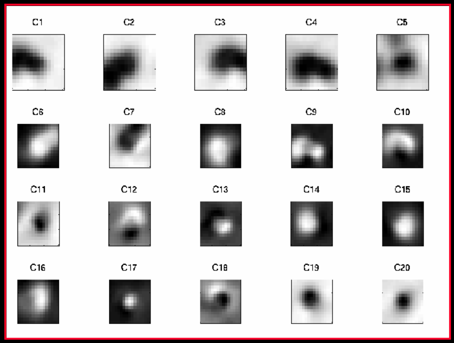
Large Data Set

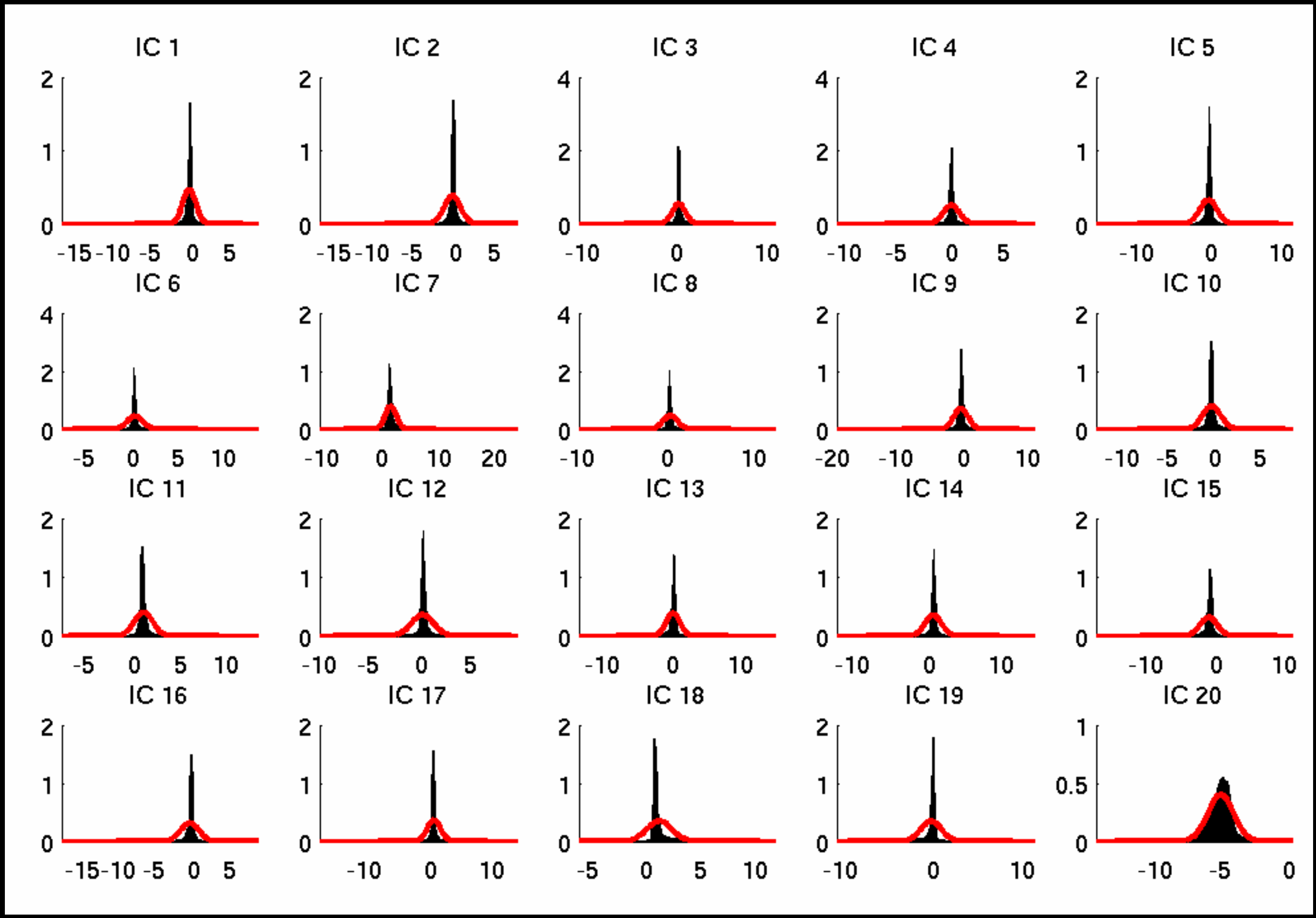
Basis Functions



- *~100 1.3 Mb images*
- *9 classes (varying dose/time treatment groups)*
- *~10 images per class*
  2 training, 8 testing
- *~10,000 15x15 subimages*
- *random sample of ~2000 training examples*
- *dimensionality reduction by PCA from 225 to 20*

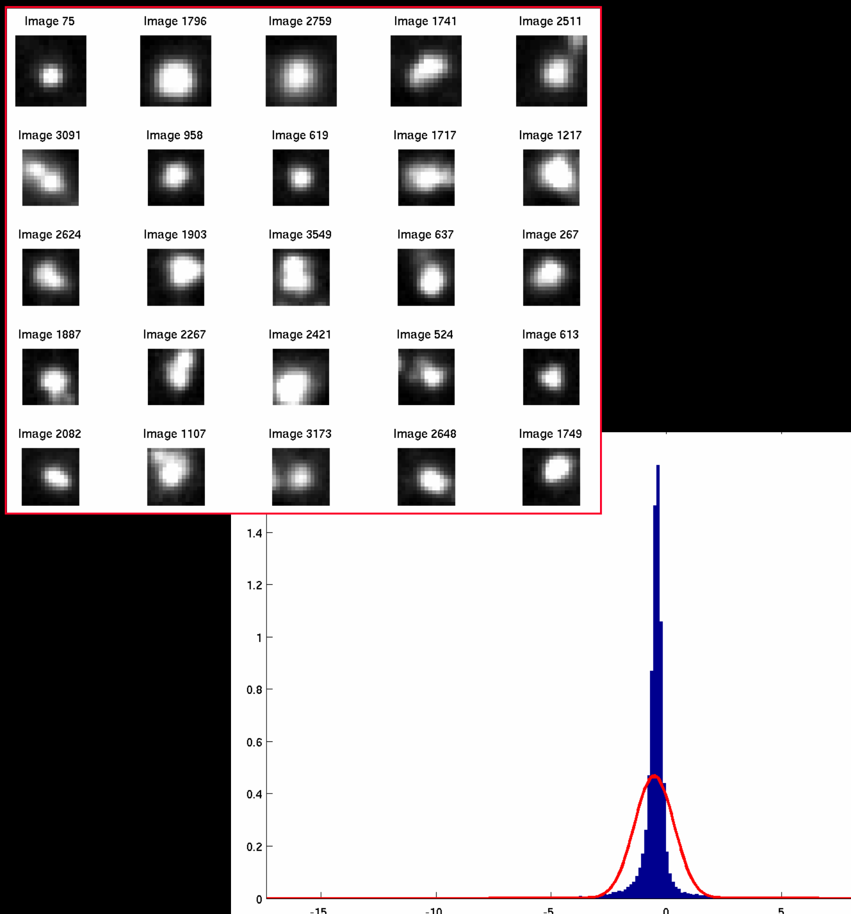
Independent Component Distributions





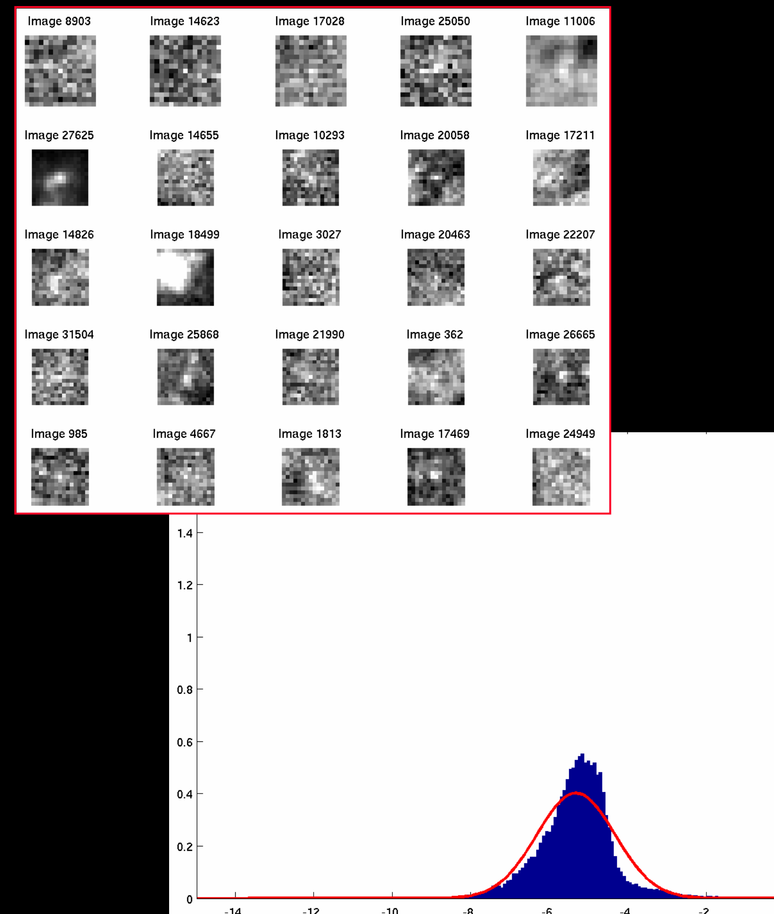
Classification by Maximum Response

Sample Subimages: FG Class



Sample IR Nucleus

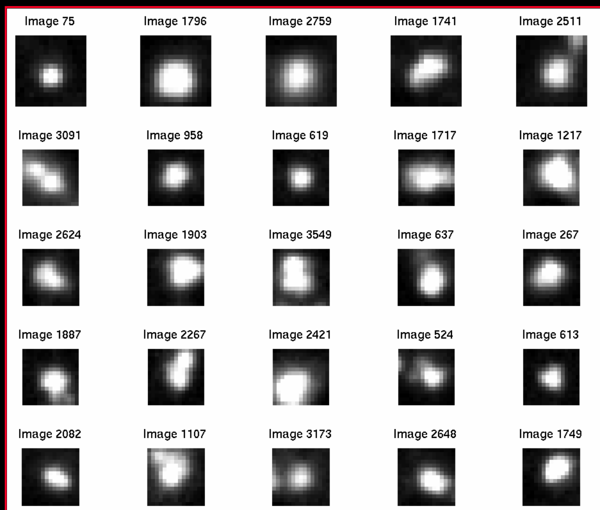
Sample Subimages: BG Class

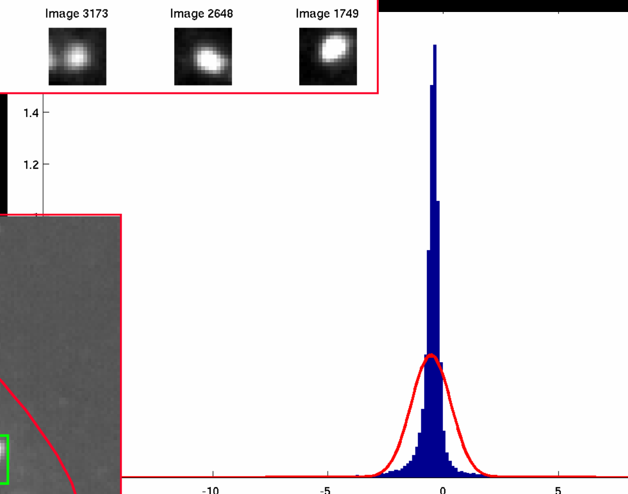


Sample Sham Nucleus

Classification by Maximum Response

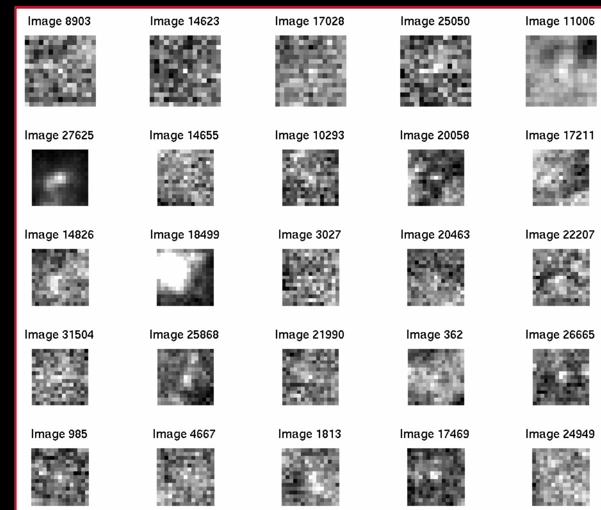
Sample Subimages: FG Class

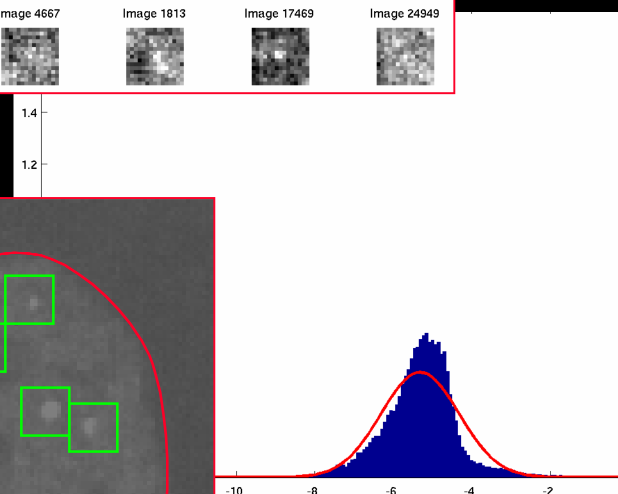


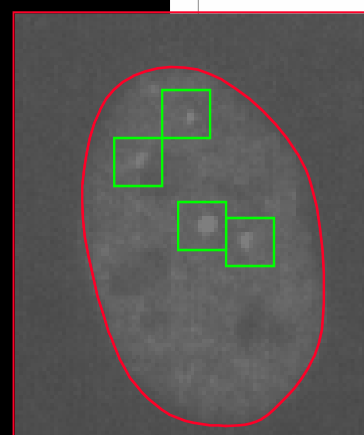


Sample IR Nucleus

Sample Subimages: BG Class

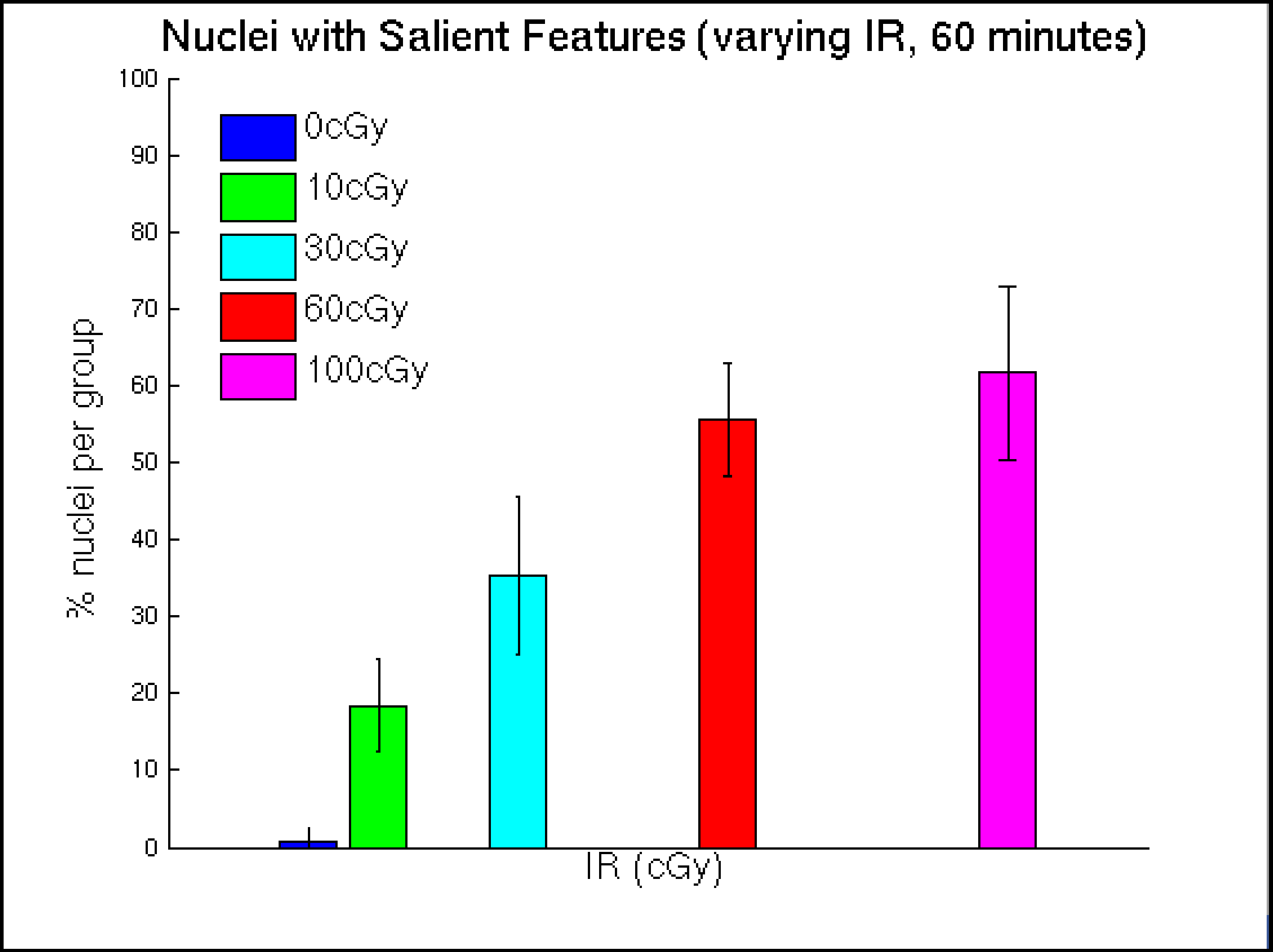






Sample Sham Nucleus

Validation


Nuclei with Salient Features (varying IR, 60 minutes)
% nuclei per group
IR (cGy)
0cGy
10cGy
30cGy
60cGy
100cGy
0
10
20
30
40
50
60
70
80
90
100

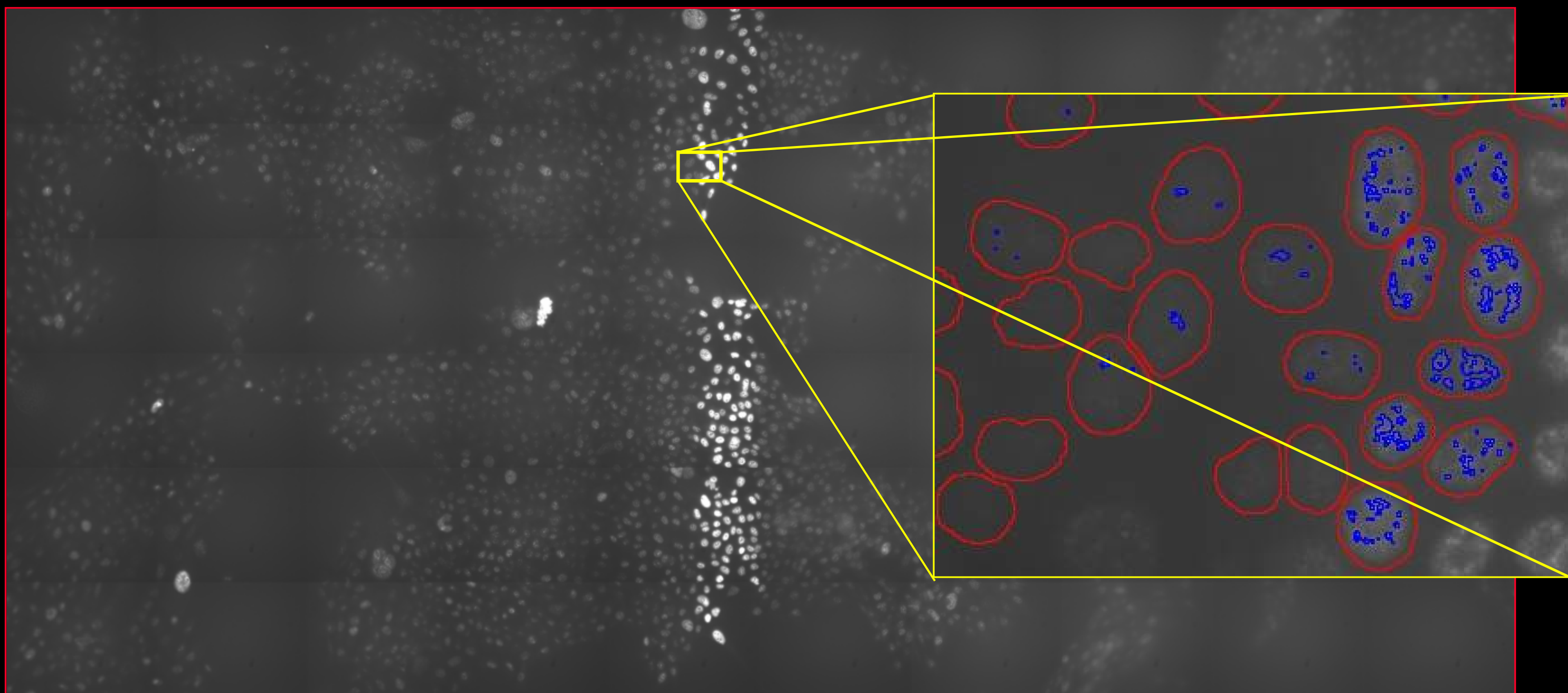
Conclusions and Current Directions

- Promising statistical method for learning representation of salient features.
- Away from hand-designed, problem-specific tools, toward general purpose solutions.

Next

- Applications to broad variety of imagery
- Feature extraction algorithms
  - subimage classification for feature detection
  - multiscale analysis
  - validation of accuracy/precision
- Cell and image classification: Bayesian modeling of treatment groups

Future Work: Bystander Effect



Thank You

Current Directions: Engineering

- Usability: trade-off between flexibility and simplicity. e.g. trainable vs. interactive vs. fully automatic
- Cross-platform implementation integrating analysis, visualization, browsing
- VTK/ITK open source commercial-grade libraries developed in medical imaging community
- Capability for distributed and parallel computing for large data sets

Feature Detection via Filter Banks

$$g(x, y, \theta, \lambda) =$$

$$\exp\left[-\frac{1}{2}\left\{\frac{(x\cos\theta + y\sin\theta)^2}{\sigma_x^2} + \frac{(-x\sin\theta + y\cos\theta)^2}{\sigma_y^2}\right\}\right]$$

$$\exp\left[\frac{2\pi(x\cos\theta + y\sin\theta)}{\lambda}i\right]$$

Parameters:

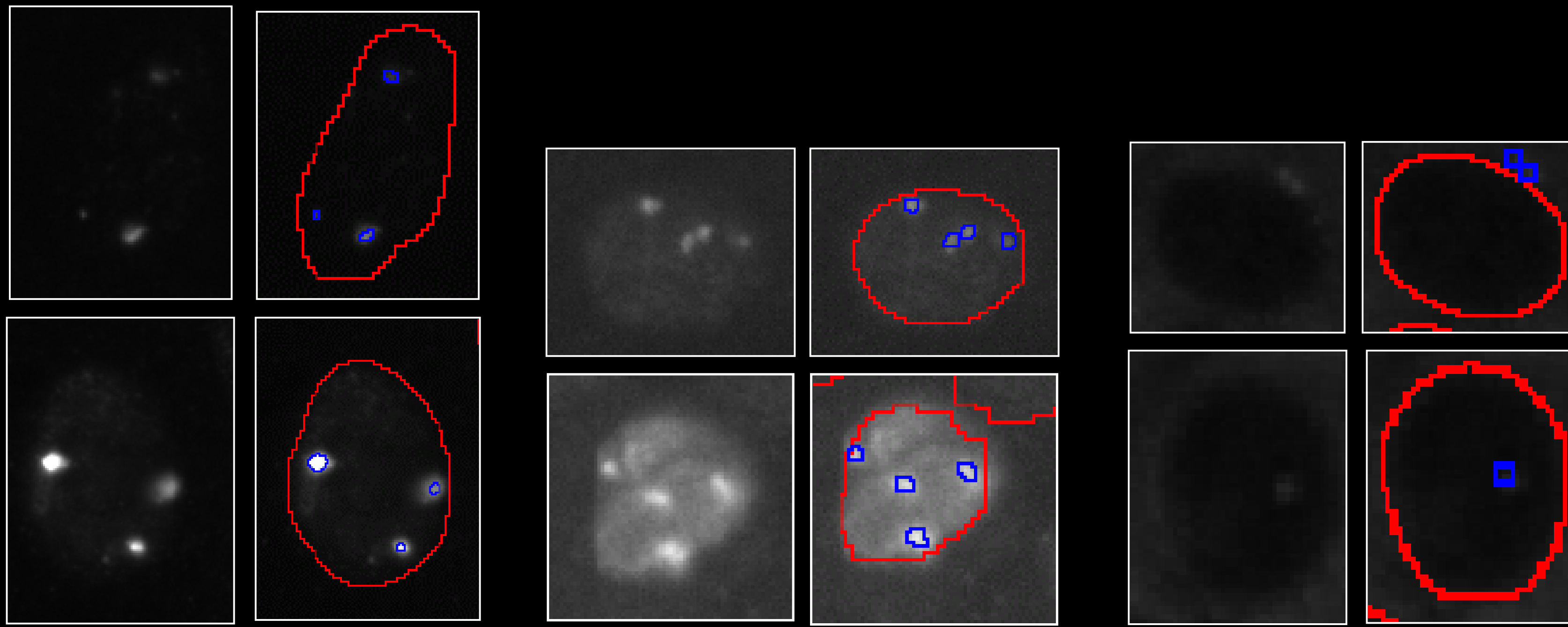
σ_x: Gaussian standard deviation in x-direction

σ_y: Gaussian standard deviation in y-direction

λ : sinusoid wavelength

θ : filter orientation

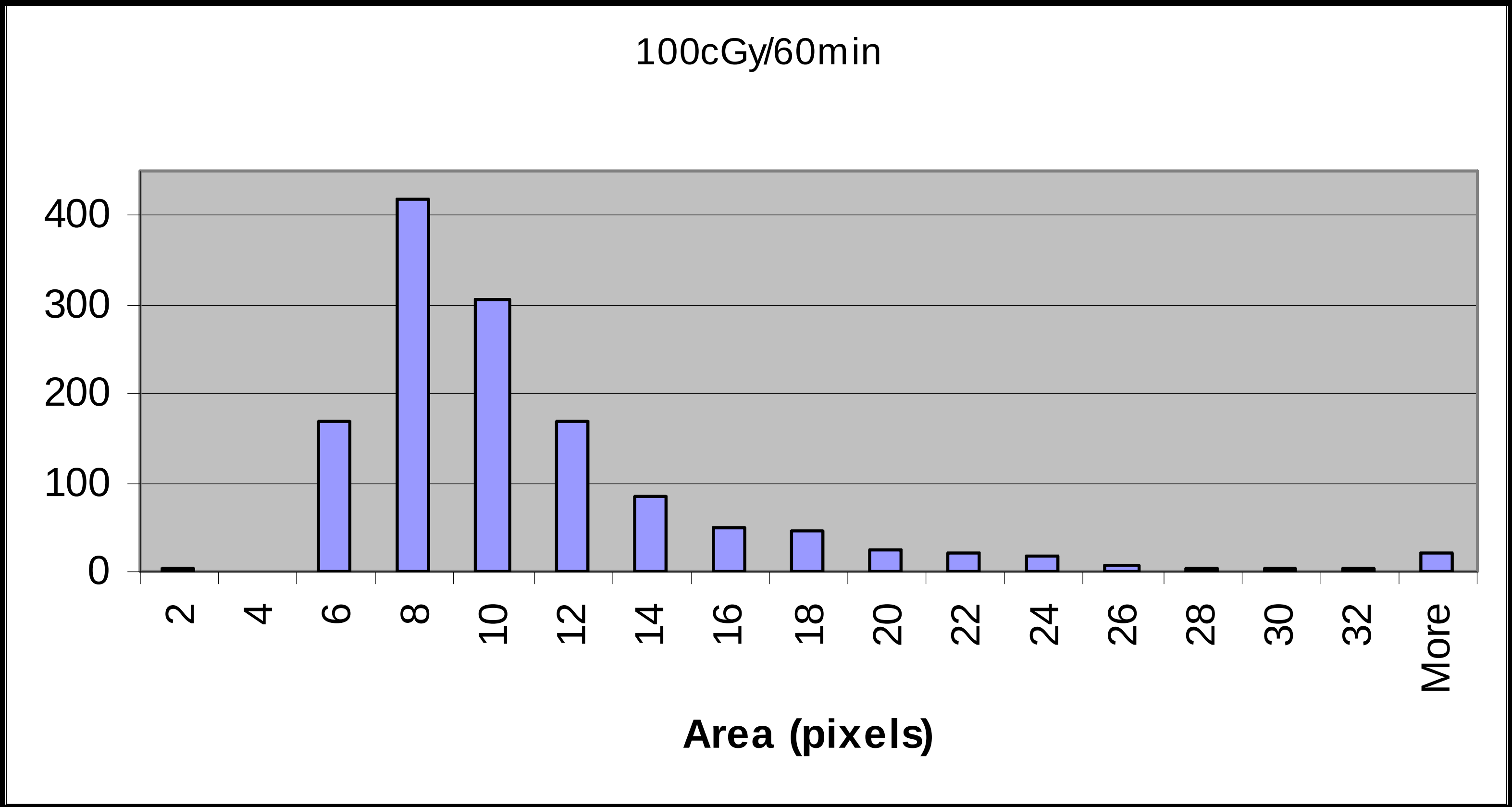
Image Processing Approach



ATMp *H2AX* *gammatubulin*

- Geometric nucleus segmentation provides local context
- Foci detection via intensity statistics and locally adaptive thresholds
- User preferences for adapting software to data set

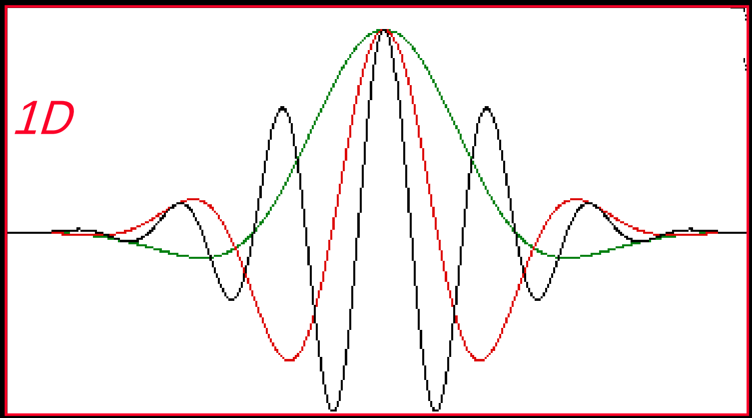
Scientific Results:
Size and Number of Foci

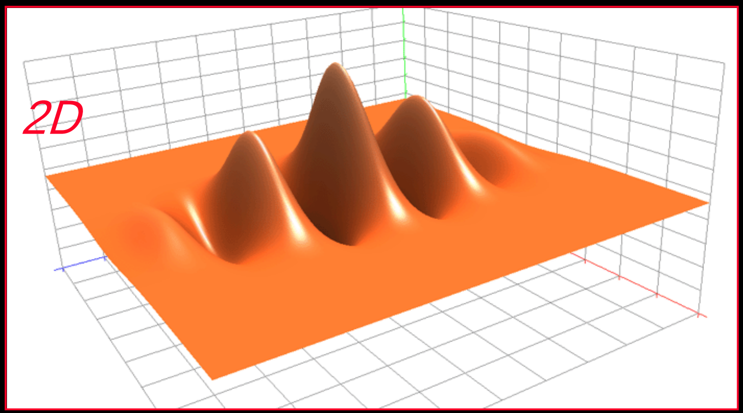


Feature Detection via Filter Banks

Gabor Wavelet Function

sinusoid modulated by Gaussian envelope





$$g(x,y,\theta,\lambda)=$$

$$\exp\left[-\frac{1}{2}\left\{\frac{(x\cos\theta+y\sin\theta)^2}{\sigma_x^2}+\frac{(-x\sin\theta+y\cos\theta)^2}{\sigma_y^2}\right\}\right]$$

$$\exp\left[\frac{2\pi(x\cos\theta+y\sin\theta)}{\lambda}i\right]$$

FastICA Implementation

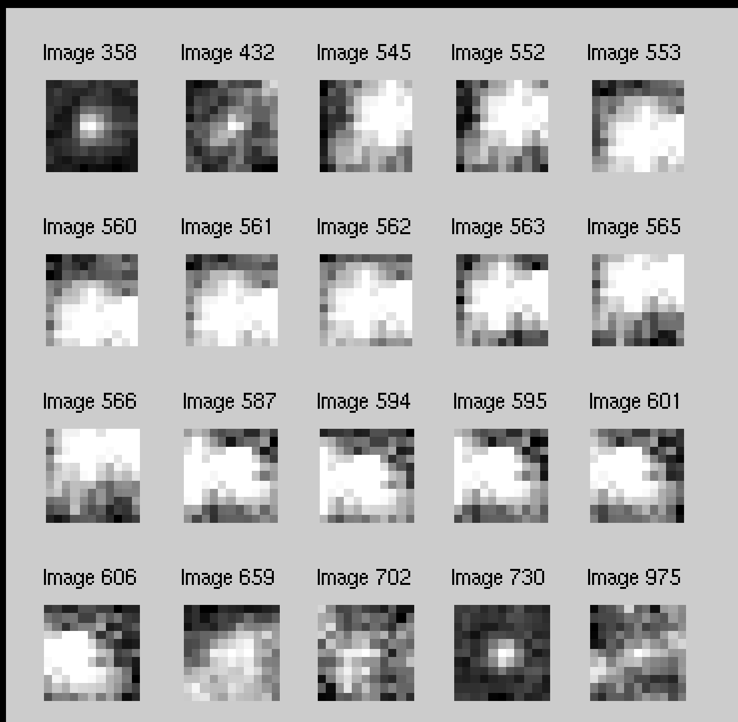
A. Hyvärinen. Fast and Robust Fixed-Point Algorithms for Independent Component Analysis. *IEEE Transactions on Neural Networks* 10(3):626-634, 1999.

- Maximizes approximation to negentropy: easy to compute, compromise between negentropy and kurtosis
- Data pre-whitened and reduced with classical PCA
- Fixed-point optimization, block-mode computation
- Parallel and distributed
- Cubic convergence rate
- c is a constant, v is a zero-mean, unit-variance Gaussian variable, G is a non-quadratic function

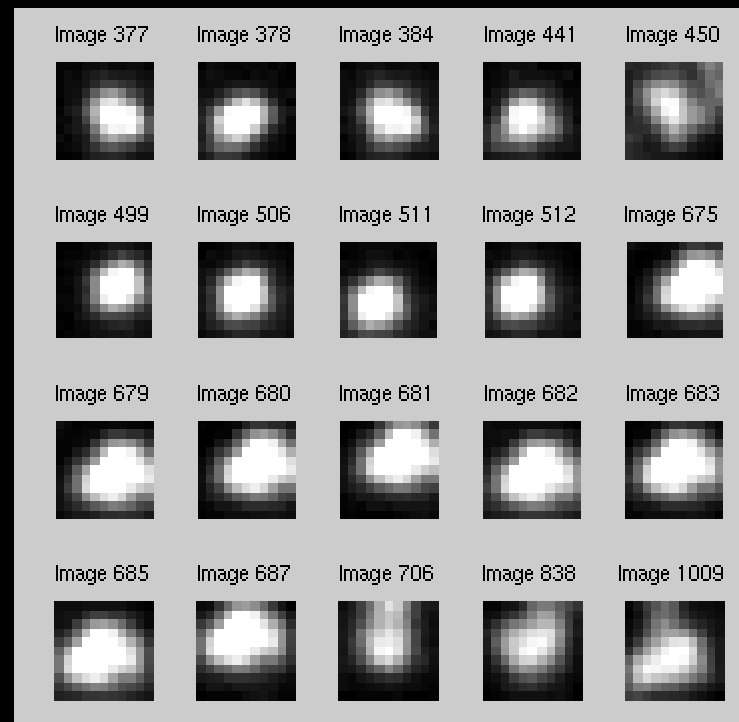
$$J(\mathbf{y}) \approx c[E\{G(\mathbf{y})\} - E\{G(\mathbf{v})\}]^2$$

Features with High Response

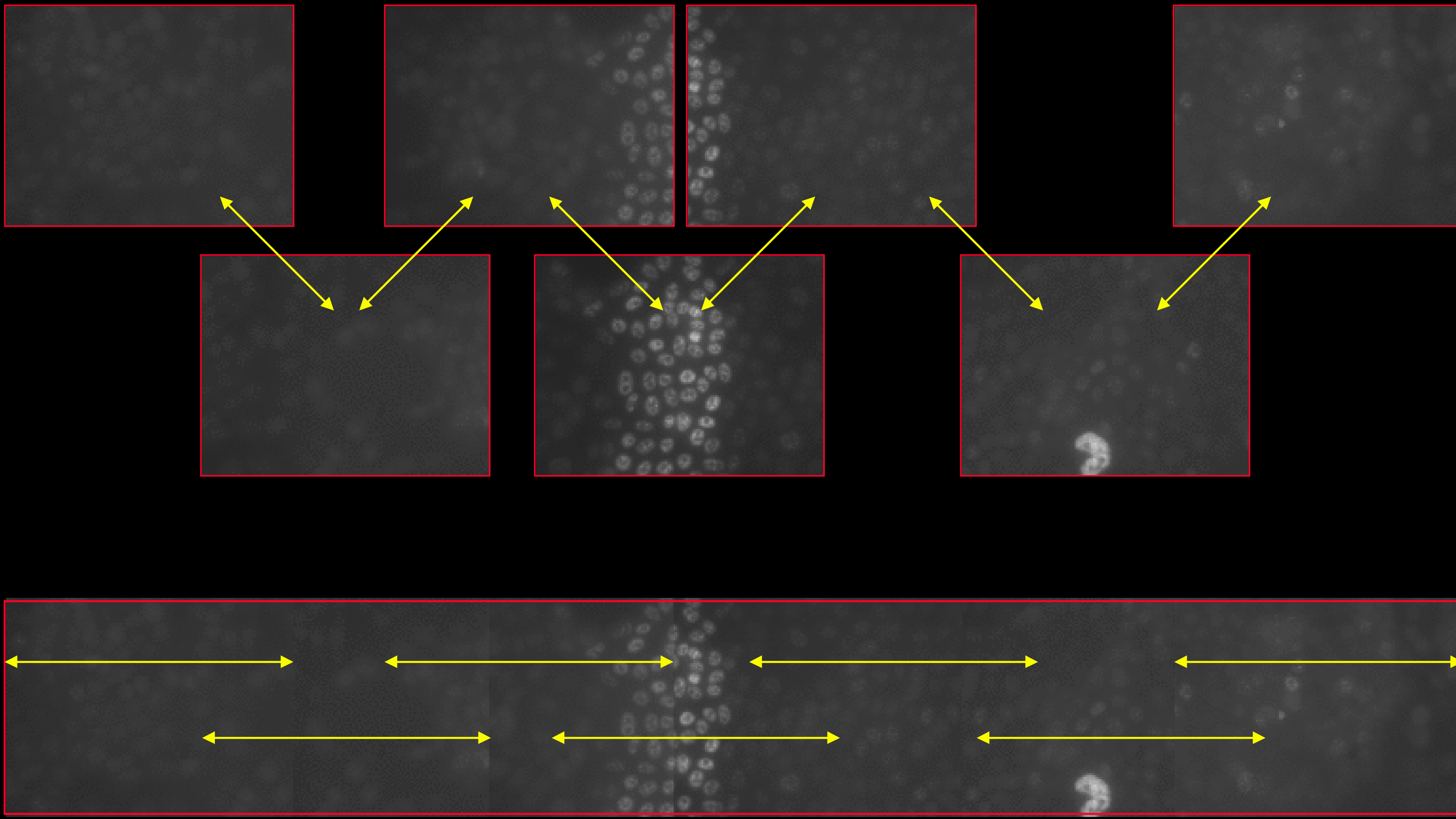
Control Group



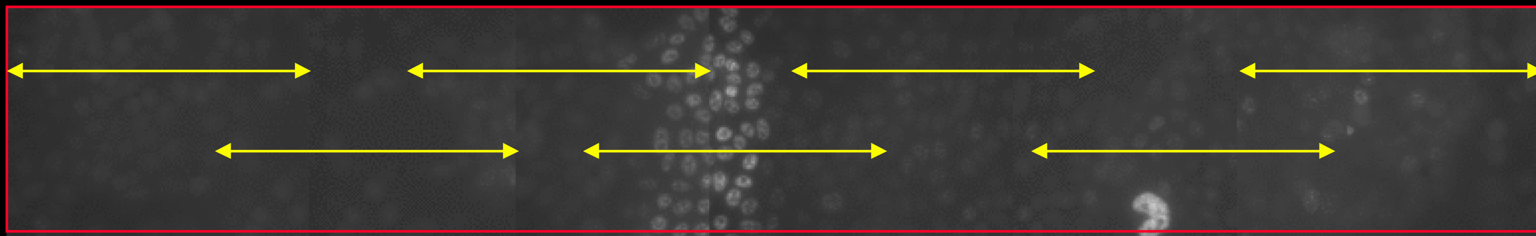
Dose: 30cGy
Time: 10 minutes

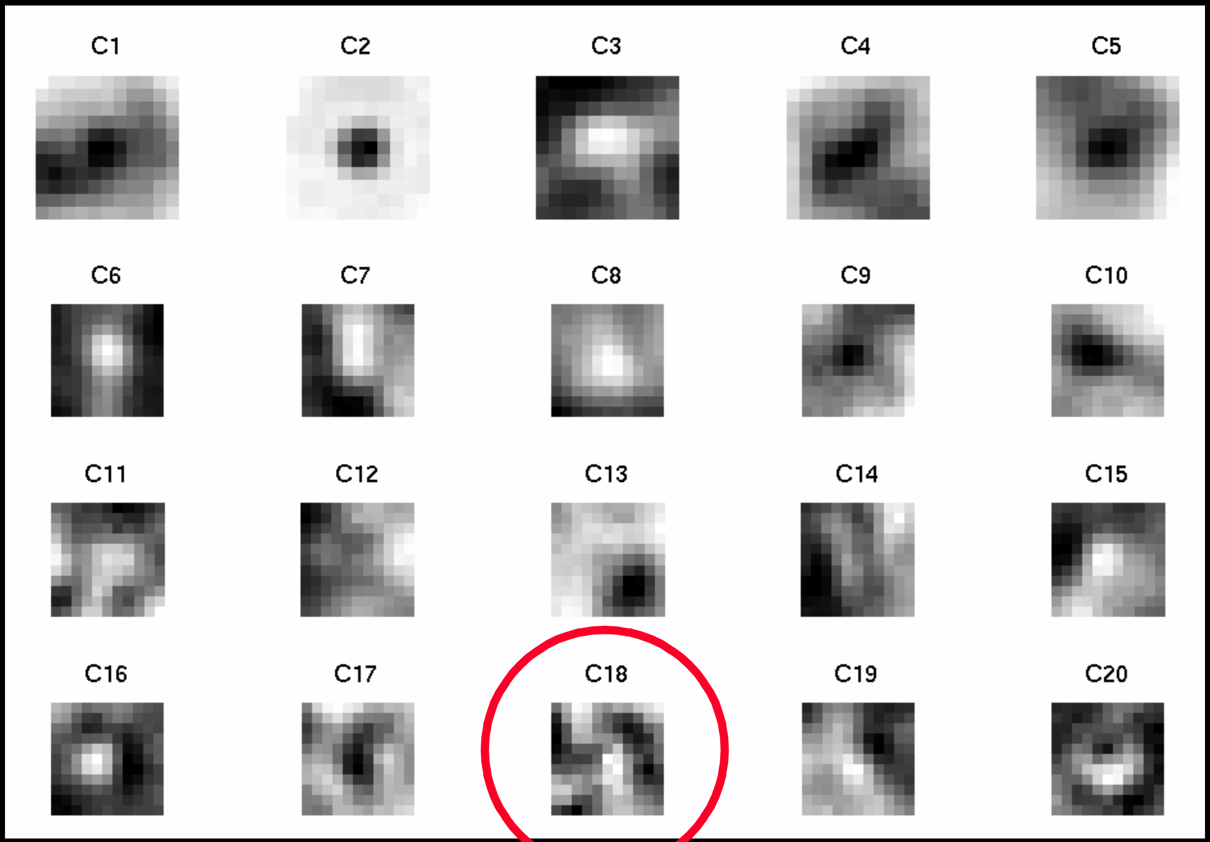


Bystander Experiment

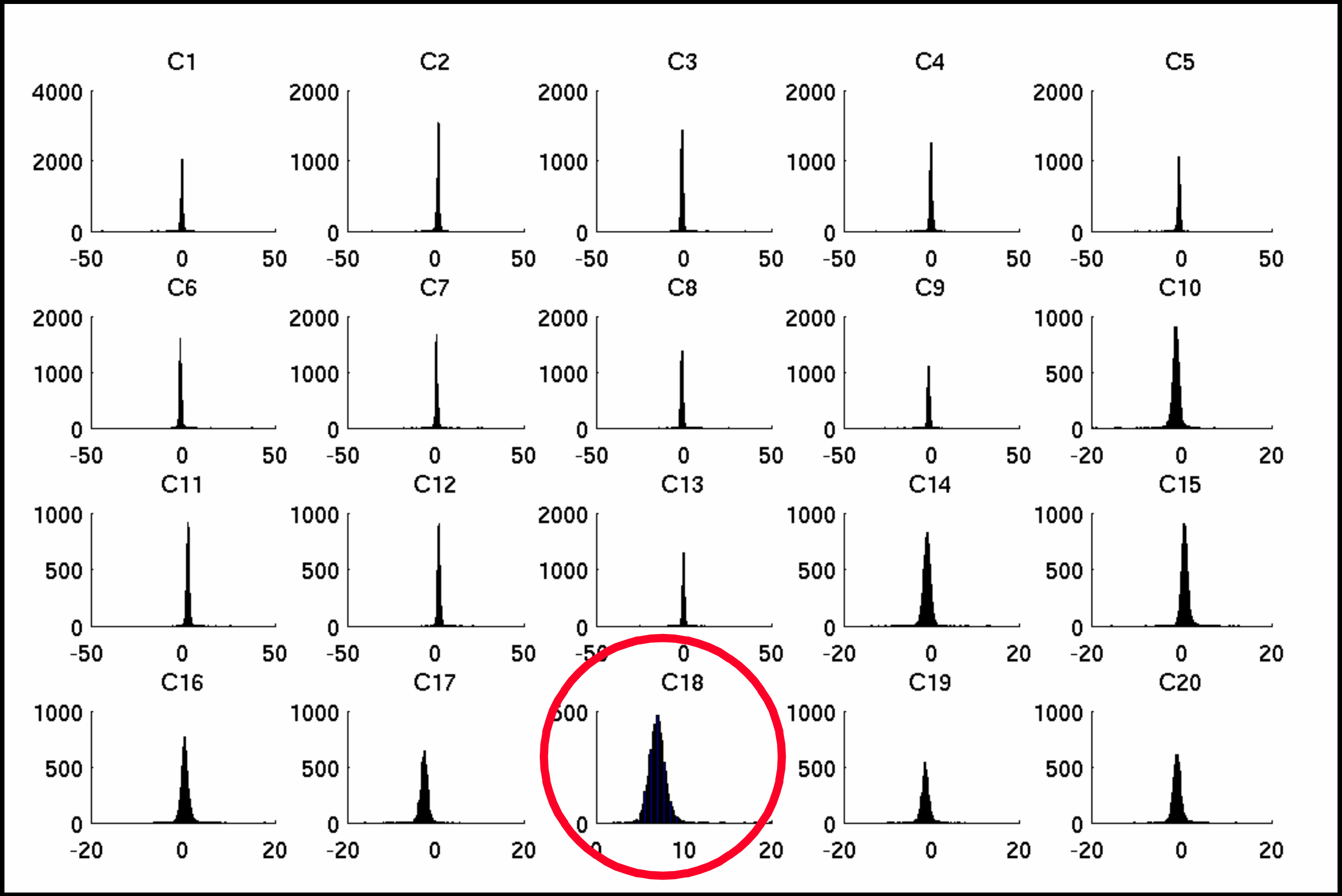


Bystander Experiment



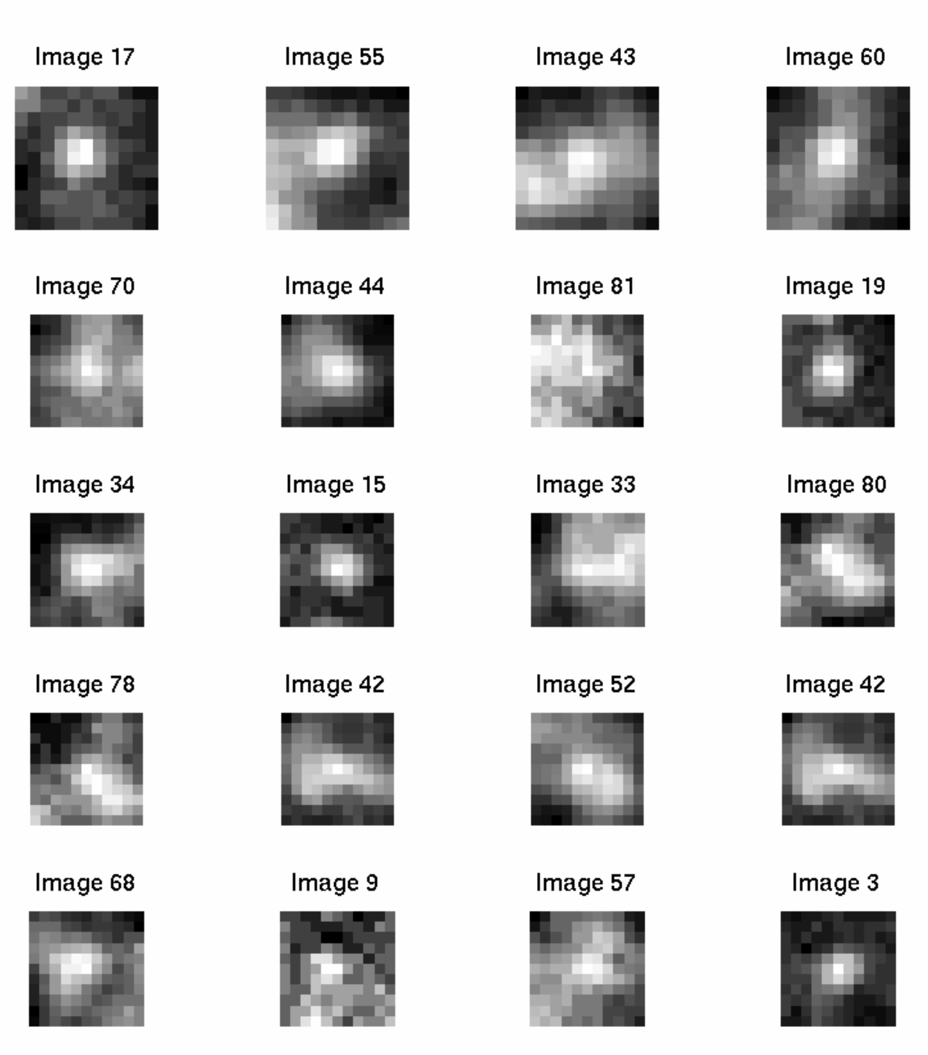


Response Histograms

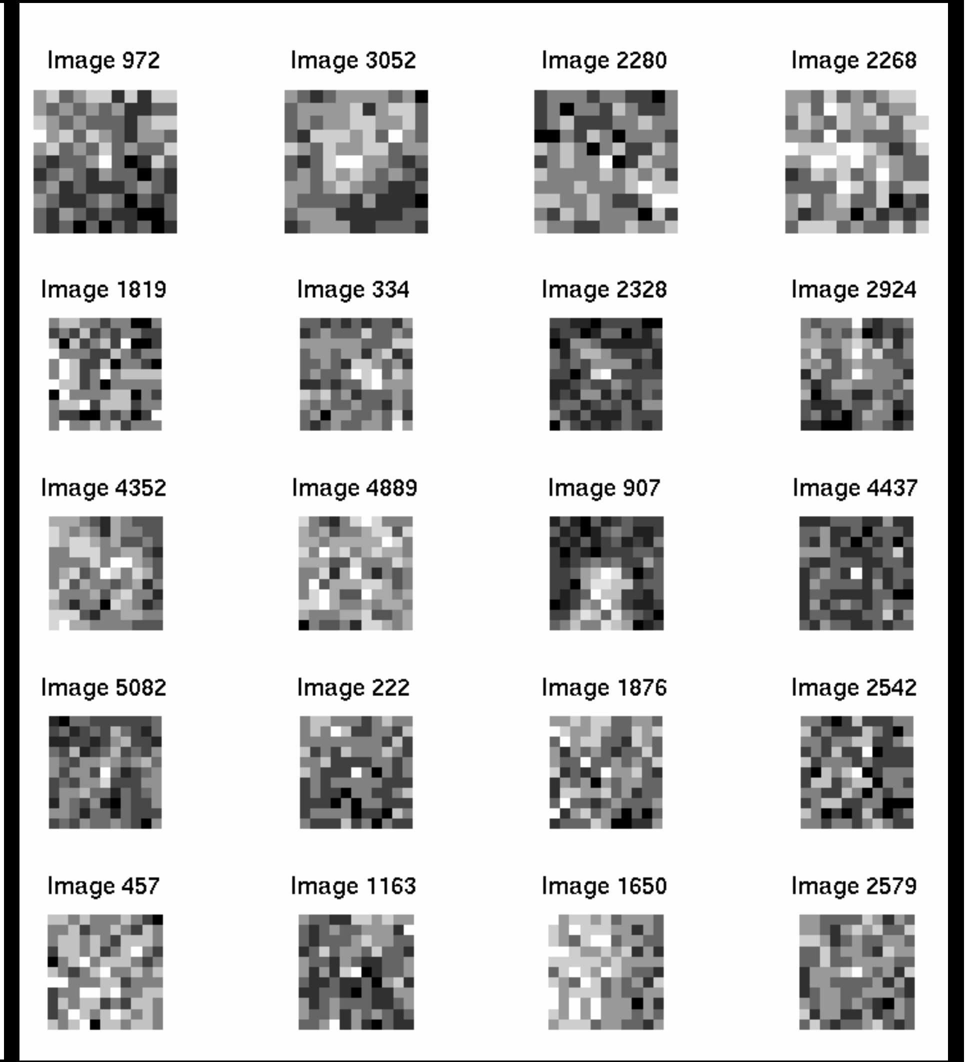


Classification by Maximum Response

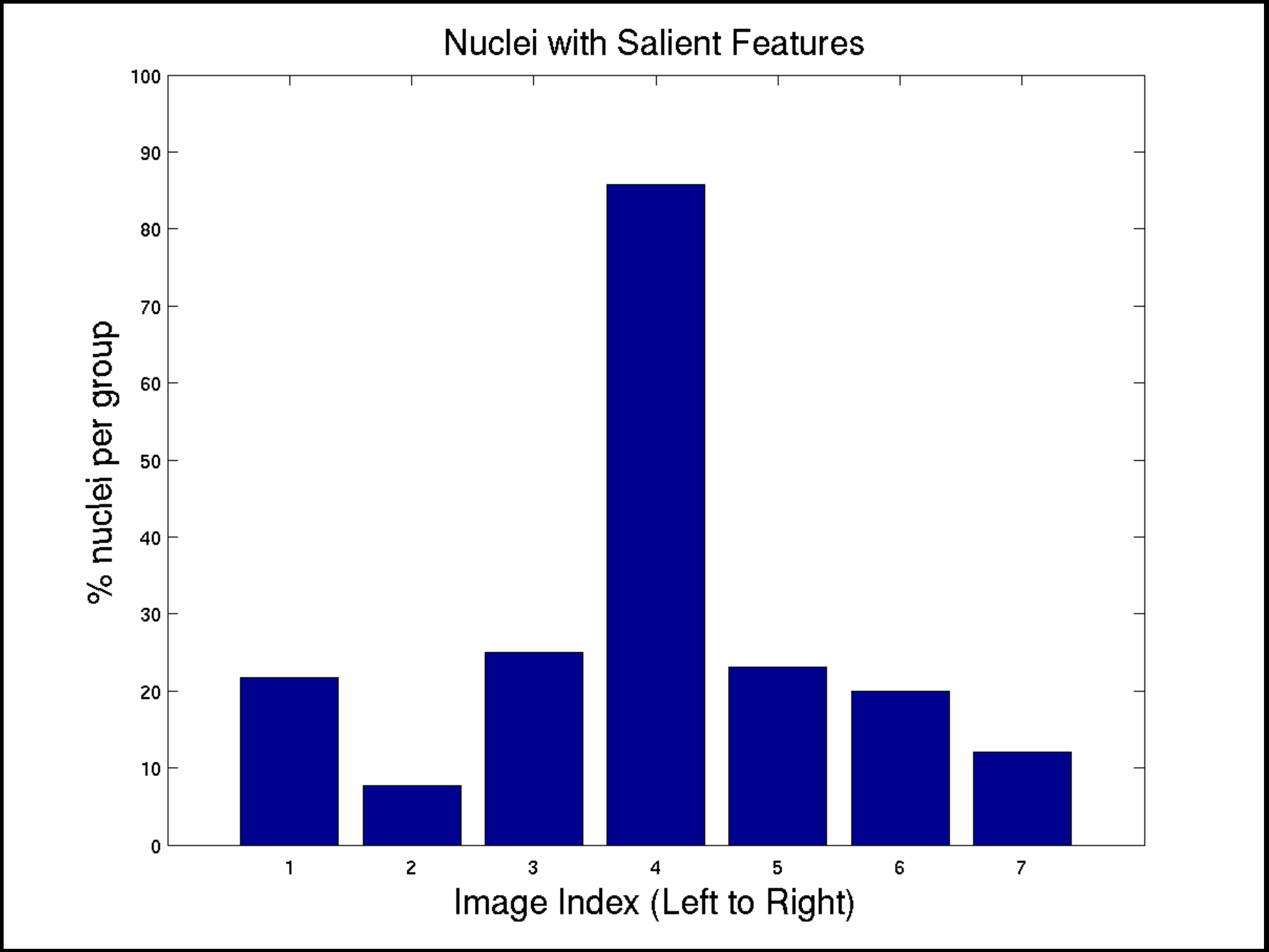
Sample Images: Class 1
(less Gaussian)



Sample Images: Class 2
(more Gaussian)



Validation

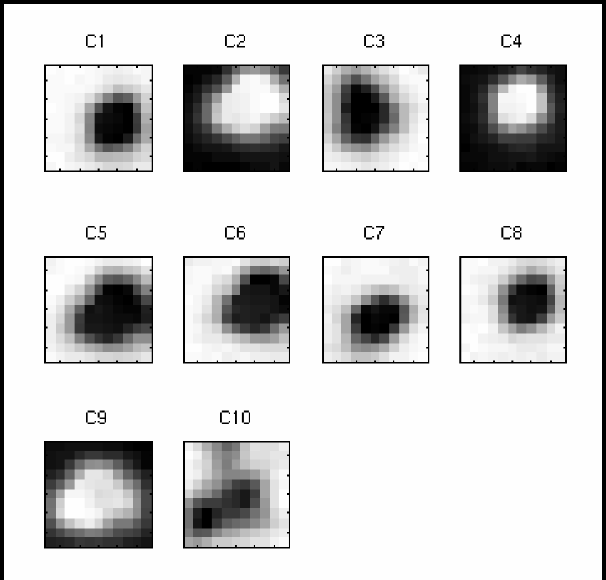


Nuclei with Salient Features
% nuclei per group
Image Index (Left to Right)
0
10
20
30
40
50
60
70
80
90
100
1
2
3
4
5
6
7

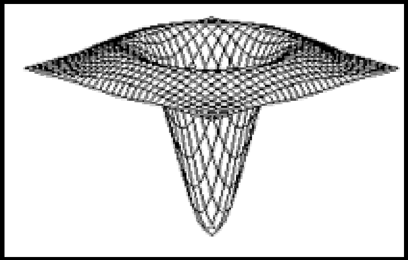
Learned Basis vs. Filter Bank

How do learned basis functions compare to typical filter bank kernels?

Learned Basis Functions



2D Laplacian of Gaussian



2D Gabor Kernel